PROTEIN COMPLEX STABILITY IN LIVING CELLS

Shannon Leigh Speer

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry.

Chapel Hill
2020

Approved by:

Gary J. Pielak

Bo Li

Zhiyue Lu

Matthew Redinbo

David Williams

ABSTRACT

Shannon L. Speer: Protein Complex Stability in Living Cells
(Under the direction of Gary J. Pielak)

Two thirds of disease-associated mutations disrupt crucial protein-protein interactions, yet these interactions are rarely studied in their native environment—in living cells. The cytoplasm of a cell is a complex and dynamic environment where the concentration of macromolecules can exceed 300 g/L. There is mounting evidence that protein stability is different in cells than it is in dilute buffered solutions, but the effects of crowding on protein-protein interactions and binding are not well characterized. While the effects of the crowded cellular environment on protein folding have been extensively studied, there is little information on how the crowded cellular environment affects protein-protein interactions. My work expands on these studies by examining the effect of the cellular environment on protein complex stability. In the first part of my dissertation, I determine how to calculate protein concentration in living *Escherichia coli* cells. Protein concentration is essential to determining the effect of the crowded cellular environment on protein-protein interactions. After determining protein concentration in living cells, I focused on controlling protein expression and determined that regular protein expression cells, BL21(DE3), cannot control protein expression, but Tuner(DE3) cells act like a rheostat and tune protein expression. Finally, I had all the pieces I needed to determine how the cytoplasm effects protein-protein interactions. My results demonstrate the importance of chemical interactions in crowding theory, contradicting

the assumption of traditional crowding theory that the more crowded environment should be more stabilizing. These results build a foundation for studying biologically relevant proteins in living cells.

To my mother, Jill Speer, for encouraging me to pursue my dreams.

ACKNOWLEDGEMENTS

I would first like to thank Gary for his guidance over the past four years. Gary has supported my growth as a scientist, mentor, and professional beyond my wildest expectations; in addition to being the most caring and sincere mentor, he has given me countless opportunities to define myself inside and outside of the laboratory and gain exposure to the biophysics community. I would not have stayed at UNC and completed my dissertation without his insight and encouragement. I am fortunate I had the opportunity to have Gary as my mentor and advisor, and I am blessed with a wonderful graduate school experience.

Second, I would like to thank my first mentors in the Pielak lab, Alex Guseman and Gerardo Perez Goncalves. Thank you both for teaching me everything I know about protein expression, purification, and thermodynamics. Like the best lab mates and mentors, I am still in touch with Alex and Gerardo and they both continue to challenge me to think deeply about my experiments and help me calm down when things go badly.

To my undergraduate research advisors and mentors at Rhodes College, thank you for your continued support after I left Rhodes to pursue my graduate degree. I would especially like to acknowledge my research advisor while at Rhodes College, Jon Russ, for reaching out to me and inviting me to do research in his lab. It is from Jon that I gained my love of analytical chemistry and quantitative experiments. I also want to

thank Dana Horgen who provided guidance when applying to graduate school and is “the person I want to be when I grow up.”

Participating in Women in Science (WinS), Women in Science Promoting Inclusion in Research Experiences (WinSPIRE) and the Graduate Business and Consulting Club has made my graduate experience at UNC truly rewarding, and I am thankful for the great students, postdocs, and PIs I have interacted with along the way. Thank you to my old lab mate, Samantha Piszkiewicz for starting the WinSPIRE program and pushing me to get involved in WinS during my graduate degree. WinS and the WinSPIRE program were instrumental to my success outside of lab.

I have enjoyed collaborating with a great many researchers during my graduate work. Candice Crilly, Joey Thole, Samantha Stadmiller, Claire Stewart, Jack Eicher, I-Te Chu, and Julia Noonan Brom have been enthusiastic about my research and have provided helpful discussion and suggestions along the way. I have never experienced a lack of encouragement throughout my graduate degree.

I had the opportunity to work with a number of talented undergraduates at UNC. Ashlee Propst has been a great friend and I have been honored to watch her grow as a scientist over the past three years as a high school student and an undergraduate at UNC. I would like to thank Francis Lauzier for his patience and understanding while I have been learning how to be a more effective teacher and mentor. Thao Nguyen, thank you for participating in the WinSPIRE program and trusting me to help you learn protein biophysics for six weeks over the summer. I am fortunate Thao and I are still in contact and she wants to pursue a research-intensive career after the completion of her undergraduate degree. I am confident she will succeed in whatever she decides to do.

It has been my pleasure to work among many great researchers in the Pielak Group, many of whom I consider both friends and colleagues. First, I-Te Chu and I quickly became close friends, and I am grateful for the many adventures we have had in and out of lab. I am positive we will remain close friends even after I leave and she finishes her degree. Claire Stewart is the most soft-spoken of the Pielak Group, but I am glad we have become friends. I am thankful for Claire's contagious positive attitude and her kind nature always picking me up after a bad day. Joey Thole, thank you for your understanding and patience when I would request to do NMR experiments certain days of the week. Candice Crilly, thank you for helping with my In-Cell NMR experiments and mastery of HSQC experiments. I would also like to thank Sam Stadmiller for her mentorship and help learning how to complete NMR experiments and for teaching me the art of In-Cell experiments. Jack Eicher made me laugh almost every day he was in lab and was always prepared with the best baked goods. And last but not least, thank you Harrison Esterly for being there to joke about conspiracy theories, lab ordering, and always being available to go get a beer.

The NMR Core staff are fabulous, and I would especially like to thank Stu Parnham, Karl Koshlap, and Greg Young for spectrometer maintenance and keeping me company in the basement and answering my NMR questions.

Outside of lab, a number of friends have helped make UNC a great place to go to graduate school. Rachel Johnson, thank you for making Genome Sciences Building the most wonderful place to work and always playing Beyonce on the speakers. Sara Maloney, my wonderful roommate and first friend at UNC, thank you for always being there with a pint of ice cream when things go wrong. I am fortunate to live with someone

who is so thoughtful, considerate and motivated. I cannot think of a better person to have spent the past 3 years living with. I would also like to thank Gina Morgan for always coming to lab with a plethora of baked goods and treats. I am grateful to all of the graduate students who have inspired me to pursue my interests and have challenged me to become a better person, including Jon Patteson, Rachel Johnson (minimus), Xiaoyan Chen, Annelise Gorensek-Benitez, Andy Chan, Erinn O'Neill, Breanne Hatfield, Katie Acken, and Adam Lescallete.

I would not be where I am today without my family, who have encouraged me to pursue my dreams and told me that anything is possible. My parents have supported me by taking numerous "panicked" phone calls throughout my graduate degree and helped me think clearly when things were tough. I love my mother, father, and sister dearly and could not have done this without them. Allison, thank you for being the best listener and twin I could have hoped for.

Lastly, I cannot imagine the past 2 years without the love of my life, Sebastian Wellford. Sebastian has been by my side throughout the ups and downs of graduate school. I look forward to spending the rest of my life with him.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

Å	angstrom
σ_{Fd}	standard deviation, fraction dimer
ΔC_{pU}	standard state, heat capacity of unfolding
$\Delta G^{\circ\prime}$	standard state, free energy
$\Delta G^{\circ\prime}_{D\rightarrow M}$	standard state, free energy of dimerization
$\Delta\Delta G^{\circ\prime}_{D\rightarrow M}$	standard state, change in the free energy of dimerization
$\Delta G^{\circ\prime}_{U}$	standard state, free energy of unfolding
$\Delta\Delta G^{\circ\prime}_{U}$	standard state, change in the free energy of unfolding
$\delta\Delta\Delta G^{\circ\prime}_{op,int}$	interaction free energy with the cellular interior
$g(L, R_d, \phi)$	work required to form a cavity of L shape, R_d radius, and ϕ volume occupancy
$\Delta H^{\circ\prime}_{U}$	standard state, enthalpy of unfolding
ϕ	volume occupancy
℃	degrees Celsius
μ_{ex}	excess chemical potential
BLIP	β-lactamase inhibitor protein
BSA	bovine serum albumin
cm	centimeter
c*	overlap concentration
CHO	Chinese hamster ovary
CRIB	Cdc42- and Rac-interactive binding
CYP	cyan fluorescent protein
Da	daltons
DEER	double electron-electron resonance

E	energy transfer
E. coli	*Escherichia coli*
ESI	Electrospray ionization
F_d	Fraction dimer
FCS	fluorescence correlation spectroscopy
FCCS	fluorescence cross-correlation spectroscopy
FPLC	fast protein liquid chromatography
FRET	Förster resonance energy transfer
FT-ICR	Fourier transform ion cyclotron resonance
GADPH	glyceraldehyde 3-phosphate dehydrogenase
GB1	B1 domain of protein G
GFP	green fluorescent protein
g/L	Grams per Liter
h	hour
H/F	proton or fluorine
HIV-Tat	human immunodeficiency virus protein Tat
HMQC	Heteronuclear Multiple Quantum Coherence
HSQC	Heteronuclear Single Quantum Coherence
Hz	Hertz
ILVA	Isoleucine, leucine, valine and alanine labeling
IPTG	isopropyl β-D-1-thiagalactopyranoside
IQGAP1	Ras GTPase-activating-like protein
IRSp53	insulin receptor substrate protein
K	temperature in kelvin
K	dissociation constant

k_a	association rate constant
k_d	dissociation rate constant
kcal/mol	Kilocalories per mole
kDa	kilodaltons
K_D	equilibrium constant of dissociation
$K_{D \rightarrow M}$	equilibrium constant of dissociation for dimerization
LB	Luria Broth
MHz	megahertz
mg	milligram
mM	millimolar
m/z	Mass over charge ratio
nm	nanometer
NMR	nuclear magnetic resonance
NmerA	N-terminal metal binding domain of mercuric ion reductase
N-WASp	neural Wiskott-Aldrich syndrome protein
MAPK	mitogen-activated protein kinase
OD_{600}	optical density at 600 nm
PDB	protein data bank
PDBID	protein data bank identifier
PEG	polyethylene glycol
PGK	phosphoglycerate kinase
pI	isoelectric point
ppm	parts per million
P_t	total protein concentration
QCI	four channel inverse cryoprobe

R_0	Förster distance
R_w	radius of water
R_a	radius of cosolute
R_b	radius of monomer
Rpm	rotations per minute
RT	room temperature times the gas constant
s	seconds
SASA	Solvent Accessible Surface Area
SPT	Scaled Particle Theory
SDS PAGE	sodium dodecyl sulfate poly acrylamide gel electrophoresis
SOD1	Superoxide Dismutase 1
$SOD^{1barrel}$	Superoxide Dismutase 1 beta-barrel structure
SOLEXY	solvent exchange spectroscopy
SW-FCCS	single wave fluorescence cross-correlation spectroscopy
TEM-1	β-lactamase protein
TMAO	trimethylamine oxide
TTHA	*Thermus thermophiles* heavy metal binding protein
$t_{1/2}q$	half-time of fluorescence recovery
μM	micromolar
XIAP	x-chromosome-linked inhibitor of apoptosis
YFP	yellow fluorescent protein

CHAPTER 1: PROTEIN-PROTEIN INTERACTIONS IN LIVING CELLS

Introduction

Two-thirds of missense mutations perturb protein-protein interactions, but there is little information on protein-protein interactions in their native environment—living cells.[2, 3] The cytoplasm of the cell is a complex and dynamic environment where the concentration of macromolecules can exceed 300 g/L and occupy 30-40% of the cellular volume, which alters the rates and equilibria of biochemical processes.[4] High concentrations of ions, metabolites, nucleic acids, and proteins define the environment where macromolecules function, and in these crowded and complex environments, proteins experience interactions with macromolecules that are absent in dilute buffered solutions where proteins are most often studied.[5, 6]

Crowding effects on protein folding have been extensively studied and are explained by a balance of excluded volume[4, 7] and soft (chemical) interactions.[7-11] Traditional crowding theory treats molecules as inert spheres that exclude volume due to steric repulsions. As a result, increased concentrations of crowding molecules will favor the more compact folded state of the protein. These steric repulsions are modulated by soft chemical interactions between a protein and the surrounding molecules. Non-specific chemical interactions can be attractive and destabilizing,[12] which favors the unfolded state of the protein, or repulsive and stabilizing, which favors the folded state.[13] Effects of crowding on protein folding have highlighted the importance

of studying processes inside the cell; however, the effects of excluded volume on more complex processes such as protein-protein association are not well-characterized.

One of the earliest analyses of protein interactions in crowded environments was in 1981 when Minton theorized that excluded volume effects would enhance macromolecular association.[7] Later in 1993, he demonstrated that crowding alters macromolecular recognition using the examples of a protein-ligand interaction and a virus binding to cell receptors.[14] In years following, investigations of crowding effects on the thermodynamics and kinetics of protein association have been conducted with the following interaction partners: calmodulin and a target peptide,[15] TEM1 and BLIP,[16] barnase and barstar,[16] and SOD and catalase.[12] Minimal effects on interactions were demonstrated in these experiments, but only polymers such as Ficoll, Dextran and polyethylene glycol (PEG) were used as crowding molecules.[16] High concentrations of polymers have been used for decades to mimic the crowded cellular environment, but recent studies have shown that polymers only have a small effect on protein-protein interactions and polymers fail to reflect the biological reality of living cells.[13, 16] Common crowding agents have little effect on protein-protein interactions; therefore, it is important to study protein-protein interactions in living cells.

Experimental determinations of interactions between proteins can be split into two types: one that views protein-protein interactions at a large scale to measure co-complex formation, and the other that views protein-protein interactions at a small scale to measure binary, direct interactions between protein pairs. For this chapter, I am going to focus on measuring the effect of the cellular environment on direct interactions between protein pairs.

Quinary Interactions

These interactions between proteins and the cellular environment can be considered quinary interactions. Linderstrøm-Lang defined the first four levels of protein structure in 1952.[17] Primary structure is the amino acid sequence. Secondary structure describes alpha-helices, beta-sheets and turns. Tertiary structure describes how secondary structure comes together to form a three-dimensional conformation of a protein, and finally, quaternary structure defines how proteins come together to form dimers, trimers, and so on. The fifth level of protein structure is quinary structure that results from weak interactions found in the crowded cellular environment (Figure 1.1).[18] Quinary structure is defined as “the transient interactions between macromolecules that provides organization and compartmentalization inside cells.”[10] Recent studies have shown that electrostatics play a major role in these types of interactions, but it would be remiss to say these are the only kinds of quinary interactions.[9, 19] Quinary interactions also include the interactions a protein experiences with small solutes and water molecules.

Figure 1.1. Five levels of protein structure by Chris A. Pielak. The tertiary and quaternary structures are based on the B1 domain of protein G (PDB ID GB1 and 2RMM). Reproduced from Cohen et al. with permission. Copyright (2016) Protein Science.[18]

Quinary interactions can impact the stability of protein complexes. Protein complexes can exist in equilibrium between a singular monomer and a dimer, and protein complex stability can be defined as the modified standard state free energy of the dimer minus that of the monomer, $\Delta G^{\circ\prime}_{D\rightarrow M}$ (Equation 1.1). $\Delta G^{\circ\prime}_{D\rightarrow M}$ equals the available thermal energy, *RT*, where *R* is the gas constant [1.987 cal/(mol K)] and *T* the absolute temperature, times the natural logarithm of the ratio of the concentrations of two states (Equation 1.2). To quantify the effect of the cellular environment, $\Delta G^{\circ\prime}_{D\rightarrow M}$ is measured in cells and the stability in buffer is subtracted to give $\Delta\Delta G^{\circ\prime}_{D\rightarrow M}$ (Equation 1.3).

$$\Delta G^{\circ\prime}_{D\rightarrow M} = \Delta G^{\circ\prime}_{D} - \Delta G^{\circ\prime}_{M} \quad (1.1)$$

$$\Delta G^{\circ\prime}_{D\rightarrow M} = -RTln(K_d) \quad (1.2)$$

$$\Delta\Delta G^{\circ\prime}_{D\rightarrow M} = \Delta G^{\circ\prime}_{D\rightarrow M,cells} - \Delta G^{\circ\prime}_{D\rightarrow M,buffer} \quad (1.3)$$

Protein Solvation

While most of the work done thus far in the field of crowding has focused on protein-crowder interactions, relatively little is still known about the effect crowding has on hydration. However, recent studies have suggested that water particles diffuse more slowly by a proteins surface and are more ordered than bulk water. Further, a simulation from the Gruebele and Pogorelov labs suggests that solutes in the cell can alter the structure of interfacial water layers and change their interaction with protein surfaces.[20] These interactions with the limited available space for a protein to interact in can change protein structure and activity.

Protein Association

The reaction describing the association of proteins A and B to form the heterodimer complex A-B is written as

$$A + B \rightleftharpoons A - B \quad (1.4)$$

For the first part of the introduction, the reaction is considered to occur at one atmosphere in dilute buffered aqueous solutions near physiological pH, i.e. modified standard state conditions. The equilibrium dissociation constant, K_D, is written in terms of molar concentrations, c, or the rate constants for formation and dissociation, k_{on}, and k_{off}, respectively, as follows:

$$K_D = \frac{C_A \cdot C_B}{C_{A-B}} = \frac{k_{off}}{k_{on}} \quad (1.5)$$

K_D has the units of concentration. The smaller its value, the more likely (i.e., the stronger) the interaction.

The most straightforward binding experiment involves fixing the A concentration and varying the B concentration. Data are often plotted as a binding isotherm (Figure 1.2) with the fraction bound (f_b) on the y-axis and the increasing concentration of monomer on the x-axis is fit to the equation below

$$f_b = \frac{[A]}{K_D+[A]} \quad (1.6)$$

to yield the dissociation constant, K_D (Figure 1.2). When the product is a homodimer (i.e., A=B) and the total protein concentration is P_t

$$f_b = \frac{\left(4P_t+K_D-\sqrt{K_D+8P_tK_D}\right)}{4P_t} \quad (1.7)$$

The dissociation constant can be converted to the free energies of dissociation ($\Delta G_D^{\circ\prime}$) using the Gibbs equation,

$$\Delta G_D^{\circ\prime} = -RTln(K_D) \quad (1.8)$$

where *R* is the gas constant, T is the absolute temperature.

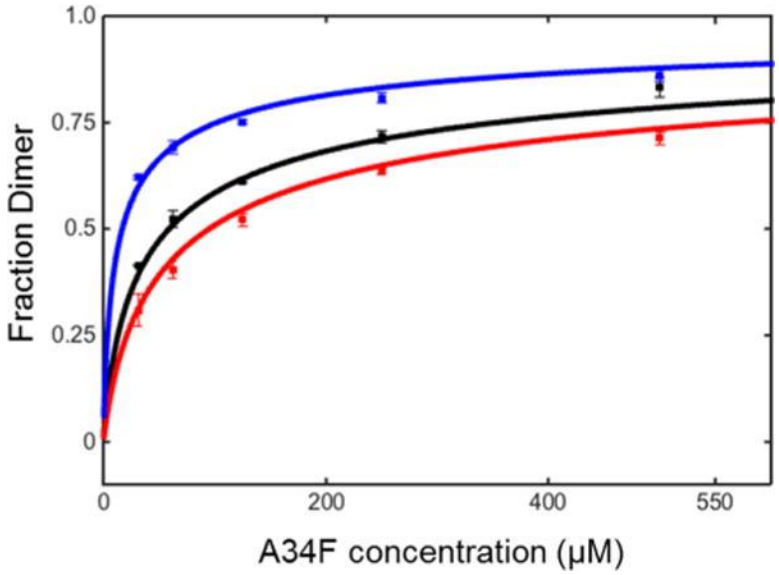

Figure 1.2. Binding isotherm for 3-fluorotyrosine labeled side-by-side dimer in solutions of cytosol (75 g/L, blue), buffer (black), and ethylene glycol (200 g/L, red) at pH 7.5, 298 K. Reproduced by Guseman et al. with permission. Copyright (2017) Biochemistry.[13]

The effect of the cellular environment on protein-protein interactions is quantified by comparing equilibrium- or rate- constants acquired in buffer alone to the constants in crowded conditions. Equilibrium constants should be written in terms of thermodynamic activity, α, of each component, *i*:

$$\alpha_i = \gamma_i C_i \quad (1.9)$$

The effect of non-ideal conditions is encoded in the unitless activity coefficient, γ. In equation 1.9, we assumed that γ is one and the activity equals the molar concentration, C, under nearly ideal conditions in dilute buffer.

Crowding changes γ, such that the K_d in dilute solution conditions must be modified by the ratio of activity coefficients to give the dissociation constant under crowded conditions.

$$K_{D,crowd} = K_D \left(\frac{\gamma_{A-B}}{\gamma_A \gamma_B}\right) \quad (1.10)$$

Thus, the crowding effects are contained in the ratio of $K_{D,\ crowd}$ to K_D. The effects of the cellular environment can then be propagated into the changes in the equilibrium binding free-energy, enthalpy, and entropy as well as activation parameters.

In what follows, we summarize the results from investigations using living cells. Values of $K_D/K_{D,crowd}$ are as large as 100, which translates to a $\Delta\Delta G_D^{\circ\prime}$ of less than 3 kcal/mol at physiological temperatures. Such changes seem small, but biological macromolecules and their complexes are stabilized by cooperative interactions, so small changes in free energy can have large biological implications. For instance, increasing the incubation temperature of alligator eggs by 4℃, corresponding to 0.01 kcal/mol of thermal energy, changes the sex of hatchlings from 100% female to 100% male.[21] We then discuss the strengths and weaknesses of the techniques used to acquire the data, interpretations of the results, and finally ideas about what needs to be brought into focus for larger picture of crowding effects.

Table 1.1. Living cells effects on protein complex stability near room temperature and neutral pH.

Protein	Complex	Cell (s)	Method	Parameters
MAPK Ste11, Ste7, Fus3/Ste5[22]	Globular/disordered	Yeast	FCCS	*K*
CDC42/N-WASp CDC42/CRIB CSC42/IRSp53[23]	Globular/globular	CHO cells	SW-FCCS	*K*
CDC42/IQGAP1[24]	Globular/globular	CHO cells and zebrafish embryos	SW-FCCS	*K*
TEM-1/BLIP[25]	Globular/globular	Hela cells	FRET	k_a,k_d, $t_{1/2}$
GAPDH/PGK[26]	Globular/globular multi-enzyme complex	U-2 OS cells	FRET microscopy/ cell volume perturbation	*K*
XIAP[27]	Disordered/molten globule homodimer	HeLa cells	DEER spectroscopy	*K*

Abbreviations: BLIP, β-lactamase inhibitor protein; CHO, Chinese hamster ovary; CRIB, Cdc42- and Rac-interactive binding; DEER, double electron-electron resonance; FCCS, FCS, fluorescence cross-correlation spectroscopy; FRET, Förster resonance energy transfer; GADPH, glyceraldehyde 3-phosphate dehydrogenase; IQGAP1, Ras GTPase-activating-like protein; IRSp53, insulin receptor substrate protein; *K*, equilibrium constant for association or dissociation; k_a, association rate constant; k_d, dissociation rate constant N-WASp, neural Wiskott-Aldrich syndrome protein; ; MAPK, mitogen-activated protein kinase; PGK, phosphoglycerate kinase; SOD, superoxide dismutase; SW-FCCS, single wavelength fluorescence cross-correlation spectroscopy; TEM-1, β-lactamase protein; $t_{1/2}$, half-time of fluorescence recovery; XIAP, X-chromosome-linked inhibitor of apoptosis

FRET

Studying and characterizing protein-protein interactions in living cells is much more challenging than studying a single protein in living cells. In the 1920s, Perrin introduced the concept of dipole-dipole interactions and the idea of distance dependence of energy transfer.[28] Fluorescence resonance energy transfer (FRET) has been used as a spectroscopic ruler to measure molecular proximity. FRET uses a higher energy donor fluorophore to transfer energy directly to a lower energy acceptor molecule, allowing for the acceptor to fluoresce. Additionally, FRET cannot be observed at distances longer than 100 angstroms, implying this is a useful technique to study protein-protein interactions in living cells. FRET is effective at a distance of 10-100 angstroms, which is equivalent to the size of macromolecules and can provide detail on protein-protein interactions. Emergence of fluorescent proteins SNAP- and CLIP- tag proteins provide FRET with the ability to monitor changes in a molecular complex in real time.[29] Förster theory states that the efficiency of energy transfer (*E*) is a function of the inverse sixth power of the distance separating the two interacting molecules (*d*) and is expressed by the following equation. Here, R_0 is represented as Förster distance, which the efficiency of energy transfer is 50%.

$$E = \frac{R_0^6}{R_0^6 + d^6} \quad (1.11)$$

FRET assays have been widely used to characterize DNA-protein, lipid-protein, and protein-protein interactions *in vitro* and in cells.

Visualizing Protein-Protein Interactions in Living Cells

Saccharomyces cerevisiae is one of the most common organisms used to study protein-protein interactions. Cyan fluorescent protein (CFP) and yellow fluorescent

protein (YFP) fusions were co-expressed containing the N-terminal transmembrane domain (NTM) of Tom70p and analyzed in living yeast cells using FRET.[30] Flow cytometry can be used in combination with FRET to make cross-talk corrections through simple compensation parameters common to all flow cytometers and can be used for examining small or large cells. Dye and Ahlquist showed that CFP and YFP can be used as indicators of protein-protein interactions (Figure 1.3).[30]

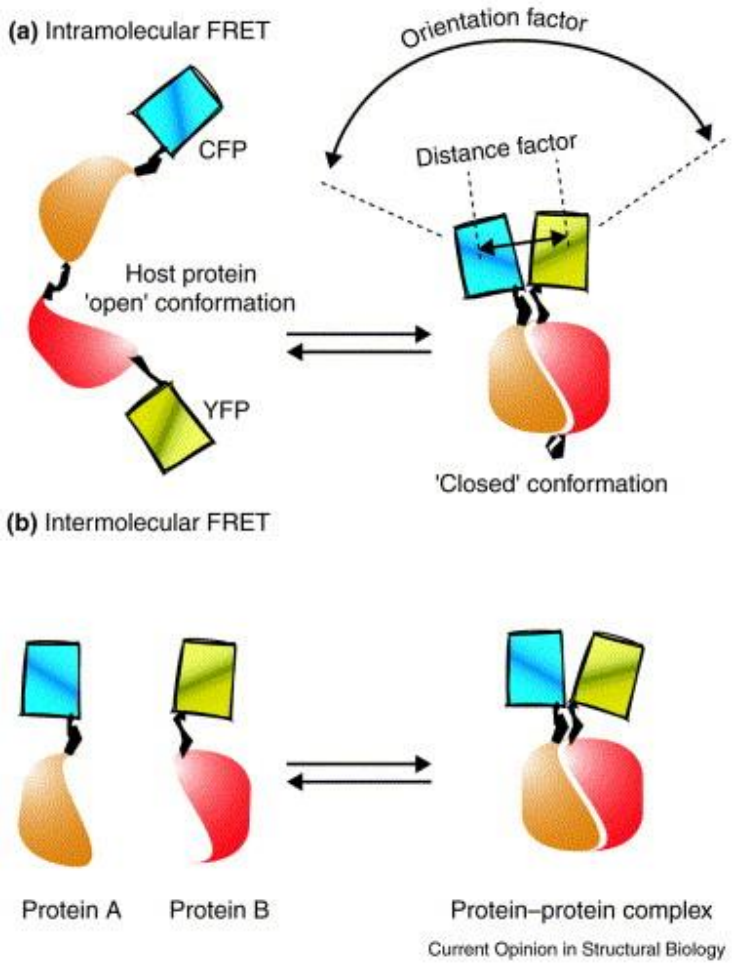

Figure 1.3. (a) Intramolecular FRET can occur when both the donor and acceptor chromophore are on the same molecule which undergoes a conformational change. In each square box corresponding to CFP or YFP (shown in cyan or yellow respectively), a diagonal line represents the chromophore. The amount of FRET transferred depends on the relative distance (<100 angstroms). (b) Intermolecular FRET can occur between one molecule (protein A) and another molecule protein (B) fused to the acceptor (YFP). When the two proteins bind, FRET occurs. When they dissociate, FRET diminishes. Reproduced from Truong et al. with permission. Copyright (2001) Current Opinion in Structural Biology.[1]

Fluorescence Correlation Spectroscopy (FCS)

Fluorescence correlation spectroscopy is a common technique for determining diffusion coefficients, chemical rate constants, molecular concentrations, fluorescence brightness, triplet state lifetimes, and other molecular parameters.[31] FCS measurement has been used to characterize protein-protein interactions because when a fluorescent ligand binds to a macromolecule, its mobility will be restricted, allowing for quantification of dissociation constants. However, the use of FCS for the study of protein-protein interactions is limited because the diffusion constant scales only to the power of one-third of the molecular mass, making it difficult to differentiate between single fluorescently labeled molecules and protein complexes.[32] To overcome the limitations of FCS, FCCS with dual-color fluorescence was created.

Fluorescence Cross-Correlation Spectroscopy (FCCS)

Fluorescence cross-correlation spectroscopy has been used to characterize and quantify protein-protein interactions *in vitro* because it allows for the detection and characterization of two protein molecules. FCCS has been derived from fluorescence correlation spectroscopy (FCS), which allows the determination of whether a protein is part of a larger complex from the diffusion rate of single molecules. Green fluorescent protein and other GFP color variants are the most commonly used fluorescent tags for FCS and FCCS. Additionally, GFP is a monomeric protein; therefore, it and its variants can be used to study protein-protein interactions.

In FCCS, interacting molecules can be studied using different fluorescent groups and the interaction can be studied by following the fluctuations in fluorescence intensity of both labeled molecules. FCCS combines the sensitivity and possibility of monitoring interactions at physiologically relevant concentrations.

DEER Spectroscopy

Double electron-electron resonance (DEER) technique, also called pulsed electron double resonance, separates pairwise couplings between electron spins from other interactions. DEER can observe dipole-dipole interactions between spin labels in two molecules even if complex formation is incomplete and transient and can thus be used to detect and quantify changes in protein-protein interactions. Similar to FRET, DEER can detect distances from 10-100 angstroms depending on protein type and environment. Further, DEER is especially useful for characterizing protein complexes that are too large for NMR or too difficult to crystalize.

Protein-complex stability in cells

Studying protein-protein interactions in living cells is complex and challenging, and most experiments have focused on fluorescence detection to assess complex formation (Table 1.1). The Knop group made the some of the first measurements of K_D in living cells using yeast.[22] Unfortunately, the dilute solution values were not reported; therefore, the results cannot be used to assess the effect of the cellular environment on protein-protein interactions.[22] Sudhaharan et al. studied the interactions between the RhoGTPase CDC42 and three of its effector proteins in Chinese hamster ovary cells.[23] Comparison with *in vitro* data shows that the cellular environment decreases the dissociation constant, K_D, by about a factor of two. The Wohland lab quantified complex formation between a CDC42 variant and an actin-binding scaffolding protein in zebra fish embryos.[24] The K_D value of 100 nM in embryos is about five-fold larger than the value determined in buffer. Phillip et al. identified a small decrease in k_a for the TEM1 β-lactamase β-lactamase inhibitor protein (BLIP) complex in HeLa cells compared to

buffer.[25] They also showed that increasing the positive charge has no effect.[25] Considering the majority of proteins in most cells are polyanions, these data also suggest a role for charge; however, the dissociation constant was not measured in HeLa cells.

The Gruebele lab assessed complex formation between glycolytic enzymes, glyceraldehyde 3-phosphate dehydrogenase (GADPH) and phosphoglycerate kinase (PGK) in human U-2 OS cells.[26] The stoichiometry changes from 1:1 in buffer to 2:1 in cells, and the K_D values were compared by taking the square root of the value determined in cells, resulting in a decrease from 20 μM to 14 μM in cells. The biggest takeaway is that quinary interactions stabilize this multienzyme complex in living cells.

The K_a of the homodimeric XIAP complex has be quantified in HeLa cells and these results demonstrated that the cellular interior destabilized this protein compared to buffer (Figure 1.4).[27] The crystal structure of the dimer shows that the protein is stabilized by a salt bridge, therefore, the cellular interior could be destabilizing to this protein-protein interaction because cells typically have a higher ionic strength it could cause the salt bridge to break. Together, these results point to the need for better quantification methods of the effects of the cellular environment in protein-protein interactions.

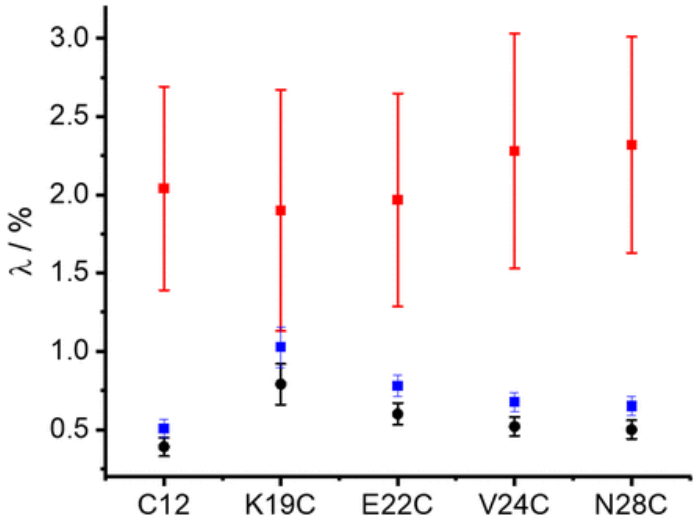

Figure 1.4. Comparison of the in-cell λ values for GdI-labeled BIR1 conjugates inside HeLa cells (black), in cell expected values (red) calculated using the solution K_D value and corrected in-cell location concentrations determined from calibration curve. The blue symbols give experiment λ values increased by 30% as described in the text. Reproduced from Yang et al. with permission. Copyright (2020) National Academy of Sciences. [27]

Methods to Quantify Protein-Protein Interactions

While most of the work done to quantify protein-protein interactions in living cells has been completed using fluorescence at physiologically relevant protein concentrations, fluorescence methods rely on large labels to quantify protein-complex formation. To offset the changes the fluorescent labels may cause for protein association it is important to compare the same system in buffer to that in living cells. One way to overcome the use of huge (up to 600 Da) labels is by using NMR to detect protein-protein interactions. NMR-based detection methods usually involve fewer structural perturbations because isotopic enrichment relies upon smaller labels like ^{19}F,[33] but NMR often lacks the sensitivity of fluorescence-based methods, so larger amounts of the protein of interest are necessary. The EPR-based DEER methodology

combines high sensitivity and small perturbations due to spin labeling, but often requires low temperatures that are not physiologically relevant.

Additionally, in-cell based measurements and dissociation constants are only apparent because of the competition between labeled protein that is required to make measurements and the unlabeled, natural version of the protein. Further, another challenge researchers have to overcome when making in-cell measurements relies on knowing the cell volume, and most of the time cell volumes are taken directly from literature values posted years ago. Even if the cell volume can be measured directly, the volume of the protein-complex is often unknown as well.

Next Steps

We would like to reiterate the importance of studying proteins in their native environment and the importance of considering hard-core repulsions and chemical interactions when studying protein-protein interactions in living cells. While there are still many additional studies needed to understand the role of quinary interactions in protein-complex formation in living cells, it is now time to study transient protein-protein interactions in living cells. The next step in understanding how protein-protein interactions occur in living cells is to understand the effect of not only charge, but also shape on protein-complex formation.

Progress has also been made in efforts to quantify protein-protein interactions in bacteria, but K_d values have not emerged yet. My graduate work has focused on quantifying the stability of a protein-protein interaction in *E. coli* cells using ^{19}F Nuclear Magnetic (NMR) spectroscopy. Our lab chose to use ^{19}F NMR to characterize protein-protein interactions because it is non-native to biomolecules, is 83% the sensitivity of

proton NMR, and simplifies the study of complex biological processes. My research has used a variant of the B1 domain of protein G, which forms a side-by-side dimer.[34] To overcome common problems with studying protein-protein interactions in living cells, first we had to determine a way to determine protein concentration in living cells,[35] and then we had to control protein expression.[36] Both methodologies have been published by our lab. In my research, I found that the intracellular environment stabilizes the side-by-side dimer and decreases the K_d from 37 μM to 11 μM in *E.coli* cells and from 40 μM to 6.5 μM in Xenopus oocytes. These findings illustrate the importance of studying protein-protein interactions in their native environment. Traditional crowding theory would predict that the more crowded cytoplasm of prokaryotic cells, 300 g/L, would be more stabilizing than the cytoplasm of eukaryotic cells, 200 g/L. However, our results from studying a variant of GB1 indicate that the eukaryotic system is more stabilizing than the prokaryotic system. This study illustrates the importance of accounting for "soft" chemical interactions when studying protein- and protein-complex stability in living cells.

Most of these studies demonstrate the importance of studying proteins in their native environment and show that when studying proteins in living cells the results often indicate little difference to buffer or stabilization.[23-26]

Summary and Closing Thoughts

While FCCS has been a great starting point for characterizing and quantifying protein-protein interactions in living cells, it is important to study protein-complex formation without the interference of large labels. DEER and NMR can overcome the limitations of FCCS and FRET.[1, 29] The use of ^{19}F NMR allows for the quantification and visualization of complex protein-protein interactions in living cells. The early results from

studying protein-complex formation in living cells indicate the importance of quinary interactions. Previously, crowding theory was predicted to mainly consist of excluded volume which always favors compact states of proteins, but studies of protein- and protein-complex stability in crowded conditions suggest that crowding is a balance of hard-core repulsions and chemical interactions. Efforts to understand these complex processes in living cells are essential for understanding cellular organization, homeostasis, and metabolism.

CHAPTER 2: CONTROLLING AND QUANTIFYING PROTEIN CONCENTRATION IN *ESCHERICHIA COLI*[1]

Introduction

Until recently, hard-core excluded volume was thought to be the key to understanding how the crowded and complex cytosol affects protein biophysics compared to dilute solutions.[3, 7] Information from in-cell protein NMR studies in *Escherichia coli* cells and lysates show, however, that chemical (aka soft) interactions between macromolecules and the protein being studied (the test protein), can be as important, or even more important, than hard-core repulsions.[37 38]To gain the most information about hard- and soft- interactions it is necessary to know the concentration of test proteins in cells. For instance, to produce the binding isotherms required to quantify the strength of protein-protein interactions using in-cell NMR, it is imperative to know, and be able to control, the concentration of the binding partners. Information about test protein concentration is also required to understand the potential for contributions from test protein-chaperone interactions.[39] However, little is known about the concentration of test proteins in cells.

We set out to control and quantify the concentration of a variant of the B1 domain of streptococcal protein G (GB1)[40] in *E. coli*. The T2Q;L5V;F30V;Y33F;A34F mutant[34, 41] was used for protein expression. We call this protein the GB1 variant.

[1] Edited from Shannon L. Speer,[1] Alex J. Guseman,[1,2] Jon B. Patteson[1], Brandie M. Ehrmann[1], and Gary J. Pielak[1,2,3,4,5*] Protein Science 28:1307-1311

Many expression systems exploit the *lac* operon, whose gratuitous inducer, isopropyl β-D-thiogalactoside (IPTG), acts in a stochastic and binary manner in commonly used *E. coli* strains, including BL21.[42] That is, expression is either 'on' or 'off'. Low IPTG concentrations induce synthesis in a small minority of cells. At high concentrations, all cells express protein at a high level. This arrangement is not appropriate for examining protein-protein interactions because a valid binding isotherm requires that every cell contains the same test protein concentration. That is, expression must be homogeneous across all cells. To overcome this problem, we use Tuner (DE3)™ cells (Novagen), which harbor a deletion (Δ*lacZY*) of lactose permease that makes every cell equally permeable to IPTG.[42, 43] The result is homogeneous expression that can be controlled, rheostat-like by varying the inducer concentration.

To understand our method (Fig. 2.1) it is useful to consider the definition of molarity: the number of moles of solute per liter of solution. In our case, the solute is the GB1 variant and the 'solution' is the cytoplasm. We use liquid-chromatography mass spectrometry (LC-MS) to quantify n, the number of moles, and flow cytometry to quantify both the number of cells, N and their average volume, V. These quantities are combined to give the concentration, C.

$$C = n/N \cdot V \quad (2.1)$$

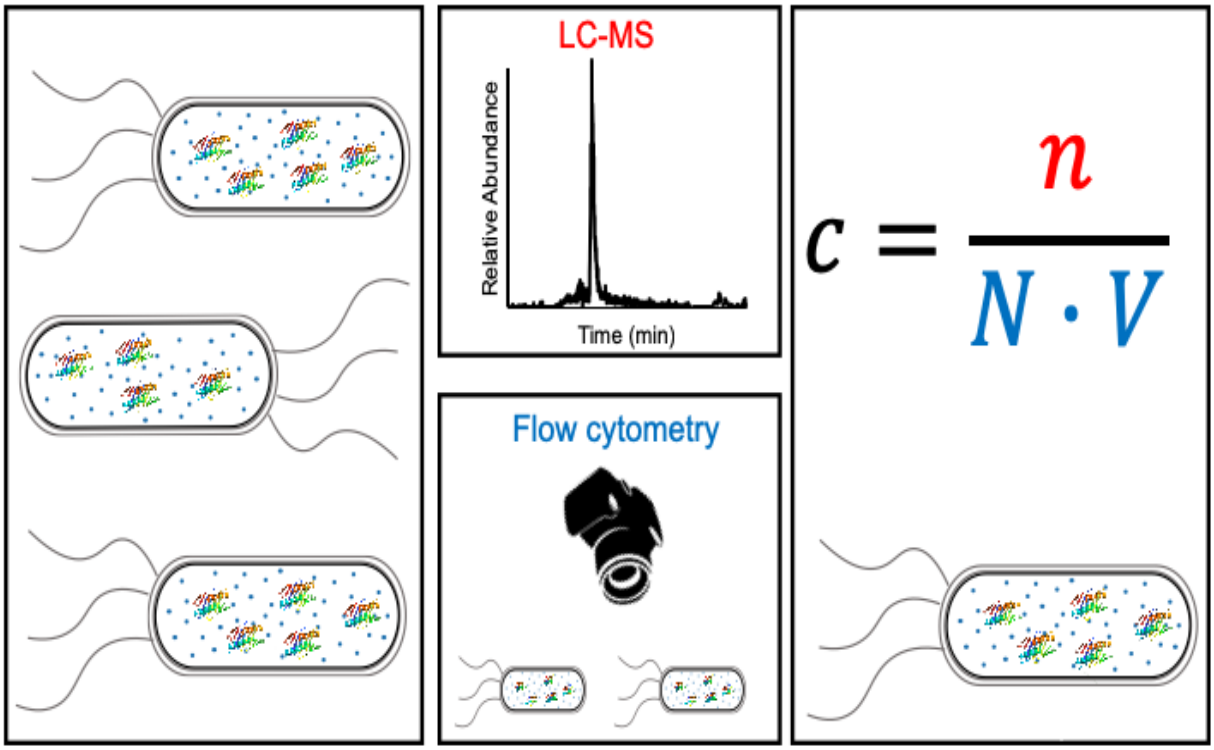

Figure 2.1. Quantifying protein concentration in cells. A sample of *Escherichia coli* cells expressing the GB1 variant (left), is subjected to flow cytometry (middle bottom) to obtain the number of cells, *N*, and the volume of cells, *V*. The cells are then lysed and subjected to LC-MS (middle top) to determine the moles of protein, *n,* and the average molarity per cell calculated (right).

Results

The IPTG concentration was varied between 0 μM and 500 μM. Chloramphenicol was used halt expression. We confirmed the expression of the GB1 variant by comparing LC-MS data from cell lysates to data from the purified GB1 variant (Fig. 2.2). To quantify the number of moles of the GB1 variant, we analyzed extracted *E. coli* lysates with LC-MS (Fig. 2.3) using the method of standard addition to overcome the matrix effect .[44]

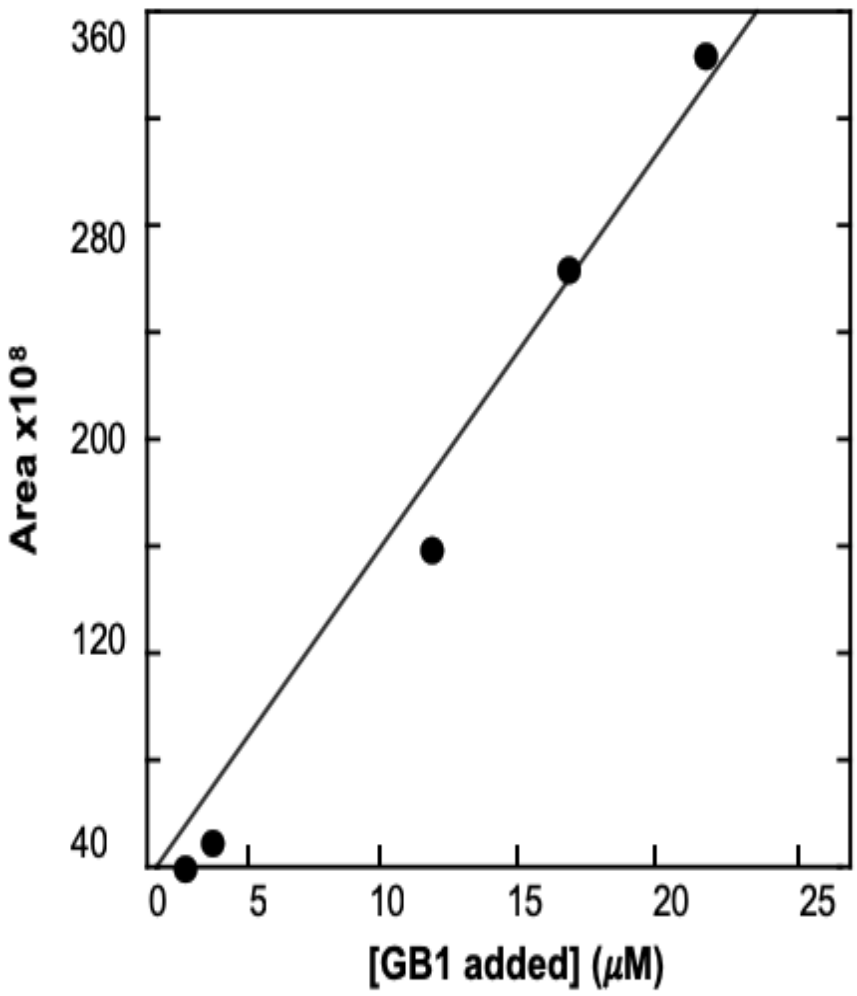

Figure 2.2. Correcting for the matrix effect. A sample of *E. coli* cells expressing the GB1 variant is lysed and purified GB1 variant is added to the cell lysate. The cells are then subjected to LC-MS to determine the moles of protein.

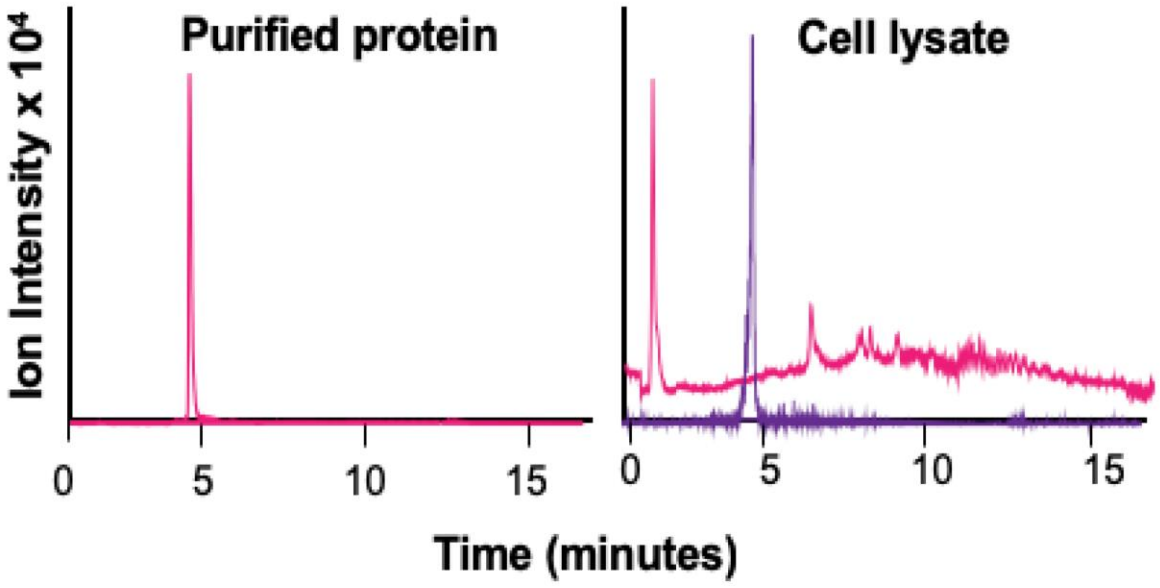

Figure 2.3. Quantifying GB1 variant concentration. Total ion chromatogram of purified GB1 (left), and a cell lysate (right, pink) with its extracted ion chromatogram (purple m/z 1040.6487 $[M+H]^+$) in *E. coli* lysate (right).

Cell size was determined from forward and side scatter of cells using flow cytometry. Our data indicate that the average length of an *E. coli* cell is 1.7 ± 0.2 μm and the average width is 1.20 ± 0.02 μm (Fig. 2.4). The volume, *V*, was calculated by assuming *E. coli* cells are cylinders. The average cell volume is 2.10×10^{-15} ± 0.03 L. This volume is larger than the predicted range of, $0.44 - 1.79 \times 10^{-15}$ L/cell,[45] because it includes the periplasm, which can amount to 20% to 40% of the total cell volume under normal growth media.[46] Accounting for the periplasm decreases the volume to $1.7 – 1.3 \times 10^{-15}$ L, consistent with predictions. The range of volumes arises from different growth conditions and acts as a reminder that analysis of cell size must be performed under the conditions used for in-cell NMR.

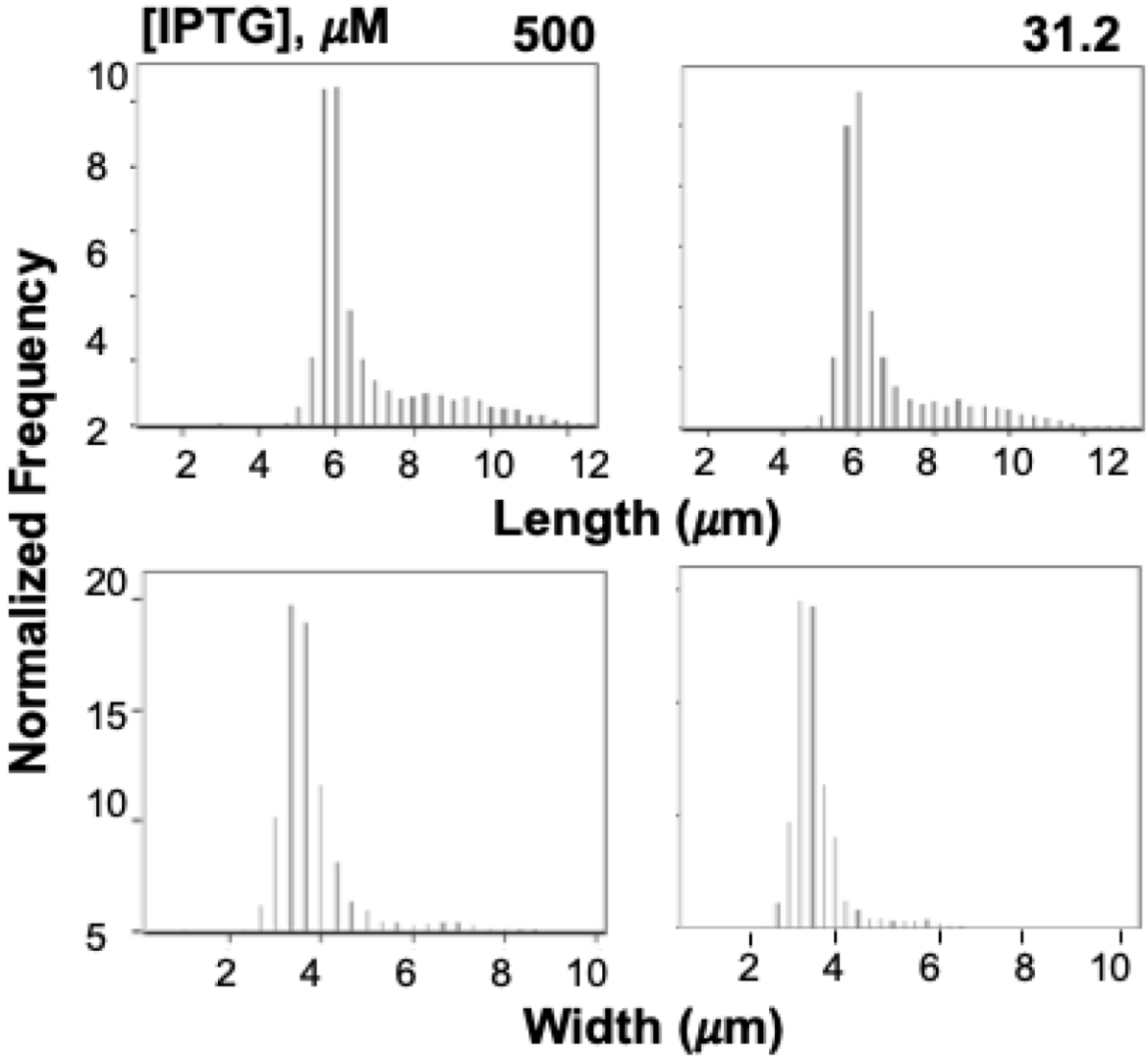

Figure 2.4. Inducer does not change the length and width of an *E. coli* cell. A sample of *E. coli* is subjected to flow cytometry and histograms of lengths and widths are obtained from forward and side scatter.

The number of cells in a sample, *N*, was determined using a counting-flow cytometer and confirmed using OD_{600}.[45, 47-51] We then quantified the average molarity per cell as a function of IPTG concentration using the number of moles of the variant, the average cell volume and the number of cells (Fig. 2.5).

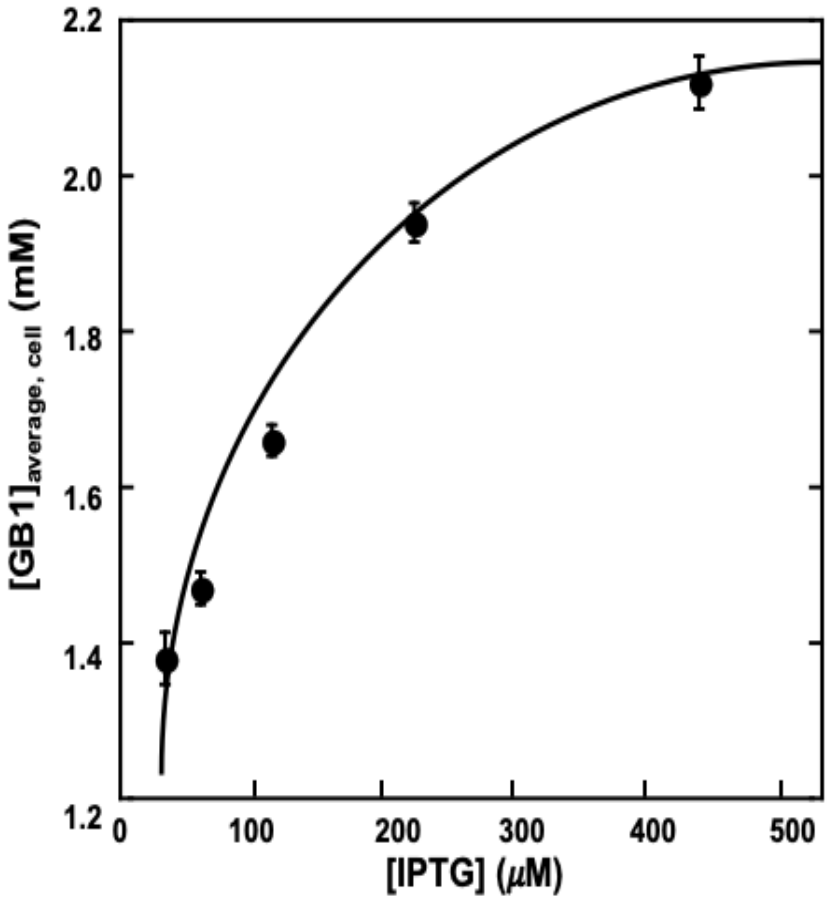

Figure 2.5. Quantifying GB1 protein concentration in Tuner(DE3) cells. Uncertainties are propagated from triplicate analysis. A sample of *E. coli* cells expressing the GB1 variant is lysed and subjected to LC-MS to determine the moles of protein and the average molarity per cell is calculated.

Discussion

The intracellular concentration of GB1 varies between 1.4 mM using 31 μM IPTG, and 2.1 mM at 500 μM IPTG (Fig. 2.5). These concentrations are larger than that of any natural protein in *E. coli*,[52] including chaperones.[39] The protein with the highest concentration in *E. coli* (~100 μM) is the chain elongation protein, EF-Tu.[52] Gro EL, the most highly expressed chaperone in *E. coli* has a concentration of less than 50 μM.[39]

What do these findings mean for using in-cell protein NMR in *E. coli* to gain physiologically-relevant data? The fact that the level of the test protein is often more than ten times the concentration of the most abundant chaperone, and the fact that experiments are limited to no more than a few hours (because of viability and leakage concerns) means that chaperone-test protein interactions are probably not important.[53]

The data also provide insight into what can be learned about protein-protein interactions in cells. The current detection limit for prokaryotic in-cell protein NMR is ~10 μM, but 100 μM is required for acquiring high-quality data, even with a cold probe.[54-57] Therefore, given the data in Fig. 2.5, dissociation constants of ~50 μM up to ~1 mM should be quantifiable.

The intracellular concentration of macromolecules under common growth conditions[58] is 200 g/L to 300 g/L[4] and does not increase when upon overexpression of a test protein.[59] Nevertheless, test-protein overexpression limits the physiological relevance of in-cell NMR data. For instance, one of the next frontiers in biophysics will be understanding quinary interactions in cells, the transient interactions between macromolecules that provide organization and compartmentalization inside cells.[10] Unfortunately, overexpression of even native *E. coli* proteins, eliminates the potential for providing detailed information about physiologically-relevant quinary interactions because overexpression necessarily spoils their stoichiometry. Such studies nonetheless provide essential information about protein biophysics in cells because the data reveal the types (charge-charge, hydrophobic, hydrogen bonding) and strengths of interactions that comprise physiologically-relevant quinary structure.[8]

Materials and Methods

A pET11a plasmid harboring the GB1 T2Q; L5V;F30V;Y33F;A34F mutant was used for protein expression. The plasmid was transformed into Tuner (DE3)™ *Escherichia coli* cells (Novagen) by heat shock. A single colony was used to inoculate 5 mL of Lenox Broth (10 g/L tryptone, 5 g/L yeast, 5 g/L NaCl) supplemented with 100 μg/mL ampicillin. This culture was incubated for 6 to 8 h at 37°C with shaking (New Brunswick Scientific Innova I26, 225 rpm), after which 500 μL was used to inoculate 50 mL of Lenox Broth and the culture shaken overnight at 37°C.

The next day, 10 mL of the culture was used to inoculate 200 mL of supplemented M9 minimal media [50 mM Na_2HPO_4, 20 mM KH_2PO_4, 9 mM NaCl, 4 g/L glucose, 1 g/L NH_4Cl, 0.1 mM $CaCl_2$, 2 mM $MgSO_4$, 10 mg/L thiamine, 10 mg/L biotin, and 150 mg/L ampicillin (pH 7.4)]. The culture was incubated at 37°C and its optical density at 600 nm (OD_{600}) was monitored (Bio-Rad Spectra Plus). Once the OD_{600} reached 0.6, protein expression was induced by adding varying concentrations of IPTG (30- to 500- μM, final concentration). After 1 h, the OD_{600} was measured and chloramphenicol (50 μg/mL final concentration) added to halt protein expression.

Aliquots of 1 mL were collected and analyzed to determine the dimensions (Amnis ImageStreamX Mark II). Calibration beads, 1 μm-diameter, were included to monitor instrument performance (Apogee Flow). Small angle forward scatter light was used as a measure of cell size. The samples were then diluted 1:1 with trypan blue to determine the total number of cells (Thermo Fisher Attune NxT).

Aliquots (1 mL) were collected and centrifuged for 10 min at 8000*g* (Eppendorf model 5430). These pellets were resuspended in 1 mL of autoclaved, deionized H_2O.

Cells were lysed by sonication (Fischer Scientific Sonic Dismembrator model 500, 15 % amplitude, 0.50 s on, and 0.5 s off for 1 min). The lysates were loaded onto solid phase exchange columns (Micro Bio-Spin 6) and centrifuged for 4 min at 1000*g*. Aliquots of the extracted lysates (75 μL aliquots) were flash frozen and lyophilized for 12 h (Labconco Freezone). GB1 standards were expressed and purified as described.[60]

The lyophilized samples were resuspended in 500 μL of LC-MS grade H_2O containing 0.01% formic acid. The resuspended samples were split into two samples, one of which was analyzed alone. Varying amounts (1 to 10 μM final concentration) of a GB1 standard were spiked into the other half of the sample. Aliquots (5 μL) were analyzed by LC-MS. Separation was achieved with a Waters CSH C18 column (1.7 μm, 150 mm x 2.1 mm) in a gradient of 0% to 100% mobile phase B over 11.25 min (initial mobile phase: H_2O, 0.01% formic acid; mobile phase B: acetonitrile, 0.1% formic acid). A Thermo Q Exactive™ HF-X Quadrapole-Orbitrap™ mass spectrometer with an electrospray ionization source operating in positive ion mode was used. The GB1 variant was detected and quantified in the cell lysate extract ([M+H] m/z 1040.6487, Fig. 3).

CHAPTER 3: RHEOSTATIC CONTROL OF PROTEIN EXPRESSION USING TUNER CELLS[2]

Introduction

Many expression systems exploit the *lac* operon, whose inducer, isopropyl β-D-thiogalactoside (IPTG), acts in a stochastic manner in commonly used *Escherichia coli* strains such as BL21 (DE3).[42] 'Stochastic' means that in such cells plasmid-driven protein expression is either "on" or "off." That is, low IPTG concentrations induce protein synthesis in a small fraction of cells, and high IPTG concentrations induce protein synthesis in all cells.

Tuner™ (DE3) *E. coli* cells are different. They contain a deletion (Δ*lacZY*) of lactose permease, which should make all cells equally permeable to IPTG[42, 43] and should result in homogenous protein expression that can be controlled, rheostat-like, by varying the IPTG concentration, but we are unaware of a direct test of this feature.

To test the potential rheostatic nature of expression in Tuner™ cells, we exploited Le Chatelier's principle; a system at equilibrium reacts to change in a way that counteracts the change. Specifically, we tested the effect of high and low IPTG concentrations on a monomer-dimer equilibrium in Tuner™ (DE3)- and BL21(DE3) cells using a variant of the B1 domain of protein G (GB1, UniProt ID P06654), whose sole tryptophan at position 43 can be easily labeled with fluorine.[33] The A34F variant of GB1

[2] Edited from I-Te Chu*, Shannon L. Speer*, and Gary J. Pielak, Biochemistry 59:733-735
*(denotes the authors contributed equally on this work)

forms a thermodynamically stable dimer.[61] The monomer and dimer exhibit unique ^{19}F chemical shifts and are in slow exchange on the chemical shift timescale.[9] The idea is that increasing GB1 concentration in cells will affect the monomer dimer equilibrium.

Materials and Methods

A pET11a plasmid containing the GB1 A34F mutant was used for protein expression.[9] The plasmid was transformed into BL21(DE3) (Novagen)- or Tuner™(DE3) (Novagen) *E. coli* cells by heat shock. A new transformation was performed every three weeks. Following overnight incubation at 37 °C, a single colony was used to inoculate 5 mL of Luria broth supplemented with 100 μg/mL ampicillin (final concentration). The culture was incubated with shaking at 37 °C at 225 rpm (New Brunswick Scientific Innova I26). After 6 - 8 h, 500 µL of the culture was used to inoculate 100 mL of supplemented M9 media. This 100 mL culture was shaken at 37 °C overnight and added to 100 mL of fresh supplemented M9 minimal media. The culture was incubated at 37 °C, and its optical density at 600 nm (OD_{600}) was monitored (Bio-Rad Spectra Plus). When the OD_{600} reached 0.45, 12 mg of 6-fluoroindole (Sigma-Aldrich) dissolved in 250 µL dimethyl sulfoxide was added.[33] The cultures were shaken for an additional 45 min, after which protein expression was induced by adding IPTG to final concentrations of 25- or 1000- µM. After 45 min, chloramphenicol (50 µg/mL final concentration) was added to halt expression.

The cells were pelleted at 2000 *g* for 20 min and washed thrice with in-cell NMR buffer [200 mM HEPES, 100 mM bis-tris propane, 150 µg/L ampicillin, and 50 µg/mL chloramphenicol dissolved in 10 % D_2O (pH_{read} 7.8)]. The pellet was resuspended in 250 µL of in-cell NMR buffer. Experiments were conducted with a Bruker Avance III HD

spectrometer equipped with a QCI cryoprobe operating at a ^{19}F Larmor frequency of 470 MHz. Spectra comprised 32768 points, 512 scans, an acquisition time of 0.9 s, and a sweep width of 20 ppm. Data were processed with TopSpin Version 3.6.1.

^{19}F NMR spectra of supernatants from cell slurries were acquired to assess GB1 leakage.[62] After each experiment, the cells were gently pelleted at 2000 g, and a spectrum of the two-fold diluted supernatant was acquired. Leakage was not observed. After the in-cell experiment, the cells were lysed, pelleted and a lysate spectrum was obtained as previously described.[5]

The protein was purified and buffer spectra acquired as described.[13] The monomer and dimer peaks were integrated to give their relative populations at five GB1 concentrations using serial dilution. The data were fit to yield a $K_{D\rightarrow M}$ at 298 K and pH 7.5 of ~20 μM.

Results and Discussion

The spectrum of A34F GB1 in buffer exhibits two peaks, one for the monomer and one for the dimer (Figure 1A). As required by Le Chatelier's principle, the fraction of dimer increases with increasing GB1 concentration (Figure 3.1A).

Next, we examined the effect of IPTG concentration on expression of the variant in Tuner™ (DE3)- and BL21 (DE3)- cells. The stochastic nature of expression in BL21 (DE3) cells is shown by the observation that the signal increases at the higher IPTG concentrations but only the dimer is observed at both concentrations (Figure 3.1B). The observation of only dimer is consistent with the fact that intracellular concentration of GB1 under these conditions is approximately 2 mM per cell.[63] In other words, GB1 is highly expressed in a small fraction of cells at low IPTG concentration and highly

expressed in all cells at a high IPTG concentration. As a control, cells were then lysed and the lysate diluted four-fold to confirm that the monomer is present (Figure 3.1B).

The rheostatic tunability of Tuner (DE3) cells is indicated by the observation that only monomer is detected at a low IPTG concentration and only dimer is detected at high IPTG concentration (Figure 3.1C). Expression increases with IPTG concentration in Tuner cells[63] and both monomer and dimer are observed at intermediate IPTG concentrations (Figure S3.1). That is, in Tuner™ cells, IPTG concentration controls GB1 concentration across all cells, and Le Chatelier's principle requires that higher intracellular GB1 concentration results in more dimer.

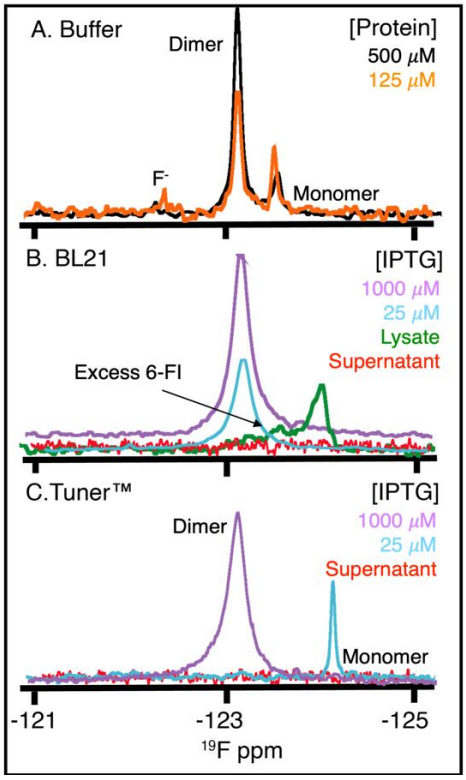

Figure 3.1. 19F NMR spectra acquired at 298 K of A34F GB1 in (A) buffer at pH 7.5, (B) in BL21- , and (C) in Tuner- cells. The resonance from F- can be seen in panel A. The resonance from excess 6-fluoroindole (6-FI) can be seen in panels B and C.

The dimer resonance has a larger width-at-half-height compared to the monomer in cells (Figure 3.1) and lysates (Figure S3.1) compared to buffer. This increase may arise from dimer-specific attractive interactions in the cellular milieu that are absent in buffer. We are testing this and other ideas *via* surface amino acid changes.

In summary, our results show that Tuner (DE3) cells can be used to vary protein expression in *E. coli* in a homogeneous, rheostatic manner. In future studies, we will quantify the equilibrium constant for dissociation of the dimer in cells and assess the effect of interactions between the dimer and the intracellular milieu.[64]

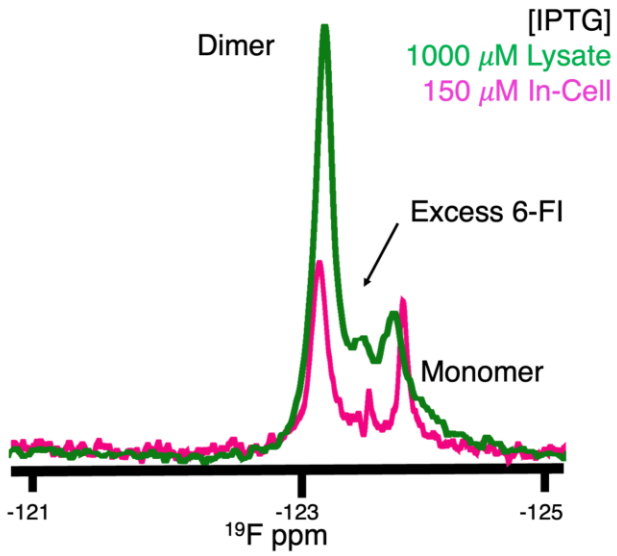

Figure S3.1. ^{19}F NMR spectra acquired at 298 K of A34F GB1 in Tuner cells at an intermediate inducer concentration concentration and a lysate spectrum at a high inducer concentration. The resonance from excess 6-fluoroindole (6-FI) is also observed.

CHAPTER 4: THE INTRACELLULAR ENVIRONMENT TUNES PROTEIN-PROTEIN INTERACTIONS[3]

Introduction

Proteins rarely work alone. Networks of protein-protein interactions turn a myriad of chemical signals into physiological responses that maintain intracellular homeostasis [65]. Therefore, it is unsurprising that nearly two thirds of disease-causing missense mutations involve protein complexes [2]. However, nearly all equilibrium thermodynamic and kinetic studies of protein-protein interactions have been performed in dilute buffer. Acquiring quantitative equilibrium data on protein complexes in cells despite the fact that living things are not at equilibrium is important for two reasons. First, the equilibrium condition is the starting point for estimating the driving force under non-equilibrium conditions [66]. Second, recent results on protein stability show that conclusions from experiments conducted in dilute solution cannot be extrapolated to the crowded and complex intracellular environment [67].

The cellular interior contains a variety of proteins, nucleic acids, and small molecules. In the bacterium *Escherichia coli*, the concentration of macromolecules can exceed 300 g/L and occupy up to 30% of volume [68]. The macromolecule concentration in eukaryotic cells is smaller [69] but still hundreds of times larger than the solute concentrations used in most *in vitro* studies. Furthermore, the majority of proteins in the

[3] Shannon L. Speer, Wenwen Zheng, Xin Jiang, I-Te Chu, Alex J. Guseman, Maili Liu, Gary J. Pielak, Conggang Li

cells studied here, *E. coli* and oocytes, are net polyanions with mean isoelectric points of 6.6 and 6.7, respectively [70]. Understanding how the intracellular environment modulates protein- and protein complex- stability is important because the totality of weak interactions in-cells form the so-called quinary structure that organizes the crowded cellular interior [10, 71, 72]. Importantly, these interactions cannot be measured in dilute solutions.

The physical consequences of macromolecular crowding arise from two components: hard-core repulsions and so called “soft” chemical interactions [64]. Hard-core repulsion is a steric effect, arising from impenetrable nature of atoms. Steric repulsions reduce the volume available to reactants and products and favor compact states. In the simple crowding theory, proteins are treated in as hard spheres. For small folded proteins such as the one studied here, stability is described by the two-state equilibrium between the biologically active folded state and the inactive, higher energy, and less compact unfolded state [73]. For protein complex formation involving folded proteins, which involves an equilibrium between constituent proteins and the complex, the complex is usually more compact [60]. Thus, hard-core effects are expected to stabilize proteins and protein complexes.

Soft interactions include charge-charge, hydrophobic and hydrogen bonds. Of these, the only repulsive interaction is that from the opposition of like charges. These repulsions add to the hard-core effect and are therefore stabilizing. The other soft interactions are attractive and are destabilizing because they favor expanded conformations that allow access to the attractive surfaces.

Until recently the focus was on hard core interactions and crowding [7] because

studies of protein stability (quantified as the free energy of unfolding, $\Delta G_U^{o\prime}$) in uncharged synthetic polymers tend to show only a stabilizing influence compared to buffer alone [64]. However, recent studies of stability under more physiologically relevant crowded conditions and in living cells show, by-and-large, a decrease in stability [67]. Studies of protein-protein interactions *in vitro* under physiologically relevant crowded conditions also show the importance of chemical interactions [60, 74].

Yeast two-hybrid techniques, co-immunoprecipitation and split systems provide simple yes/no, information about protein interactions in cells, however, these methodologies are also prone to false positives. The protein-protein interface can also be characterized by in-cell NMR. However, quantification of the equilibrium thermodynamics of complex stability is challenging in cells. The few reports tend to involve heterodimerization, which can be complicated to assess because the concentration of both reactants must be controlled, most FRET labels can add hundreds to thousands of Da must be included and because of competition between the labeled proteins, which are required for quantification, and the natural, unlabeled, proteins already present in cells [11, 12, 75-79].

Here, we use the simplest type of complex and a label that adds only a few Da. The homodimer formed by the A34F variant of the 6.2 kDa domain of protein G (GB1, Fig. 4.1A) is well characterized in buffer [61] and under crowded conditions [13, 60]. Both the monomer and dimer are folded, and dimerization involves neither a significant conformational change nor a large reduction in surface area. The effect of surface charge has also been assessed [9]. Importantly, GB1 is not found in the cells, which means there are no competing specific interactions from endogenous proteins.

To gain broadly applicable information, we assessed complex stability in a prokaryote (*E. coli*) and a eukaryote (*Xenopus laevis* oocytes). To detect dimerization, we used simple fluorine labeling techniques and ^{19}F NMR, which is advantageous for studying sensitive living cells because acquisition of spectra requires only a few minutes. Replacing a side chain hydrogen atom with a fluorine adds only 18 Da and has a small or negligible effect on structure and stability. Furthermore, the reaction between the monomers and the dimer occurs much more slowly than the frequency difference between their resonances, which means that monomer and dimer concentrations are directly proportion to the area of their resonances [61]. Specifically, the sole tryptophan of GB1 was replaced by 6-fluorotryptophan (6FW) [33] or its three tyrosines were replaced by 3-fluorotyrsine (3FY) [80]. These two labeling strategies were used because they provided the highest quality data in the two different cell types.

To quantify dimer formation in dilute solution, we measured the fraction of GB1 dimer as obtained from integration of the ^{19}F resonance, as a function of GB1 concentration. These binding isotherms (Fig. 4.1C and 4.2C) were fit to equation 1 [81], where $F_d$is the fraction of dimer and P_t is the total GB1 concentration in cells or in buffer alone to yield the equilibrium dissociation constant, $K_{D\rightarrow M}$ (fig. 4.S1 and 4.S2).

$$F_d = \frac{4P_t + K_{D\rightarrow M} - \sqrt{{K_{D\rightarrow M}}^2 + 8P_t K_{D\rightarrow M}}}{4P_t} \quad (4.1)$$

Stability is defined as the free energy of homodimer dissociation, $\Delta G^{o\prime}_{D\rightarrow M} = -RTln(K_{D\rightarrow M})$, where *R* is the gas constant in kcal/mol and *T,* is the absolute temperature. Uncertainties are represented as the standard deviation of mean from triplicate measurements. The stability in buffer alone was determined as described [82].

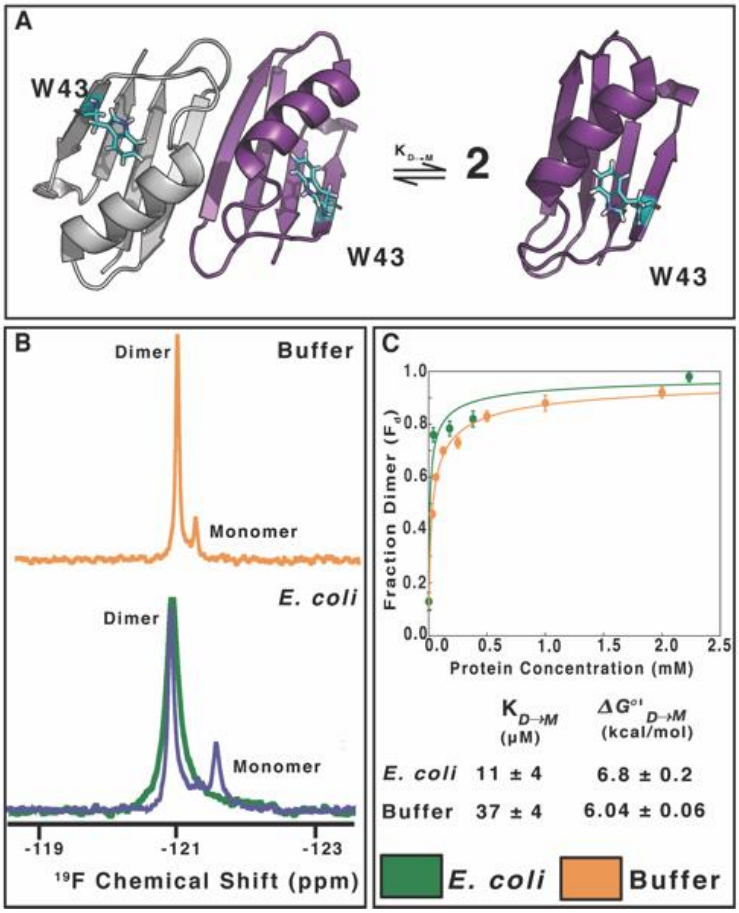

Figure 4.1. Complex formation in *E. coli*. (A) Dissociation of the A34F GB1 side-by-side homodimer (PDB ID code 2RMM) showing the 6-fluorotryptophan at position 43. (B) ^{19}F NMR spectra acquired at 298 K of the 6-fluorotryptophan-labeled protein in dilute solution 0.500 mM GB1 in 20 mM phosphate buffer (pH 7.5) and *E. coli* using inducer concentrations of 1.0 mM (green) and 0.24 mM (purple) with the cytosol buffered at pH 7.6. (C) Dissociation constants were quantified from the binding isotherms acquired in cells (green) and buffer (orange). Error bars represent the standard deviation of the mean from triplicate analysis.

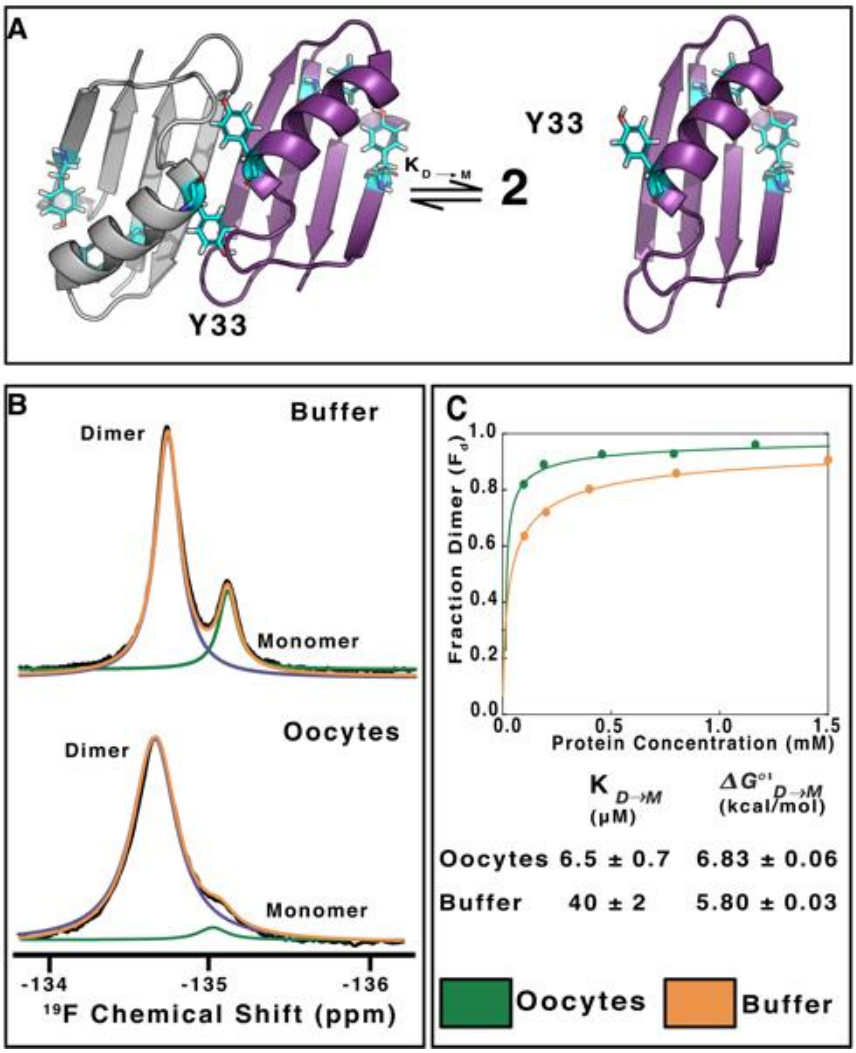

Figure 4.2. Complex formation in *X. laevis* oocytes. (A) Dissociation of the A34F GB1 side-by-side homodimer (PDB ID code 2RMM) showing the 3-fluorotyrosines at positions 3, 33, and 45. (B) ^{19}F NMR spectra acquired at 288 K of 3-fluorotyrosine-labeled protein in 20 mM phosphate buffer (pH 7.4) and oocytes (black, as acquired; purple, deconvoluted dimer; green, deconvoluted monomer; orange, sum of deconvolutions) (C) Dissociation constants were quantified from binding isotherms acquired in oocytes (green) and buffer (orange). The uncertainties, which are from least squares fitting, are smaller than the points.

To assess dimerization in *E. coli* (Fig. 4. 1A), GB1-expressing plasmids were transformed into a strain containing the ΔlacZY mutation to assure tunable control of protein expression GB1 [83], the protein was labeled by adding 6-fluoroindole to the

medium [60]. Leakage was tested [62] after each experiment. None was observed [83] (fig. S4.3). The ^{19}F resonances from the monomer and dimer (Fig. 4.1B) are broader in cells than in buffer (Fig. 4.1B) due to attractive interactions between the GB1 and other cytosolic proteins and slightly increased viscosity in cells (fig. S4.4) [80]. Quantification of populations requires complete recovery of longitudinal magnetization, which was accomplished with a 4 s delay between acquisitions. Monomer and dimer resonances were fully resolved, and only a few Hz of line broadening was applied prior to Fourier transformation. The GB1 concentration was varied by using different inducer concentrations (fig. S4.5). Mass spectrometry combined with flow cytometry was used to assess P_t [84]. The NMR and protein concentration data were combined to construct binding isotherms (Fig. 4.1C), which were used to determine $\Delta G^{o\prime}_{D\rightarrow M}$ by least square fitting.

For *X. laevis* oocytes experiments, we microinjected the purified ^{15}N-enriched, 3FY-labeled A34F protein (Fig. 4.2A) along with a proteinase inhibitor cocktail to prevent degradation, and recorded ^{19}F spectra (Fig. 4.2B) and ^{1}H-^{15}N HSQC spectra (fig. S4.6). The internal pH of oocytes is 7.4 [85]. The similarity of spectra acquired in oocytes to those acquired in buffer alone at the same pH (fig. S4.6) suggests that the structures are unchanged. As noted [86], resonances are broader in oocytes than in buffer, but not as broad as in *E. coli*. After each in-cell NMR experiment, the supernatant was reexamined by NMR. No leakage was observed (fig. S4.7). ^{19}F spectra acquired before and after acquisition are nearly identical (fig. S4.7), indicating that the dimer and monomer concentration is constant in oocytes throughout the experiment. We used a combination of NMR and microscopy to determine P_t (fig. S4.8).

The concentration of GB1 in the NMR tube was obtained by reference to a sample of known concentration in buffer alone.[69] We used a microscope to measure an average oocyte volume of 1.0 ± 0.2 μL (fig. S4.8), consistent with another report [87]. P_t is the product of average volume and the number of oocytes used in the NMR experiment. ^{19}F spectra were fitted to Lorentzian line shapes before Fourier transformation (Fig. 4.2B). The combined results used to construct binding isotherms (Fig. 4.2C), which were used to determine $\Delta G^{o\prime}_{D\rightarrow M}$ by least square fitting. A complete list of $K_{D\rightarrow M}$ and $\Delta G^{o\prime}_{D\rightarrow M}$ values acquired in *E. coli*, oocytes and buffer is given in table S1.

The interior of both *E. coli* cells and oocytes stabilize the A34F dimer relative to buffer (Figs. 4.1C and 4.2C). Such stabilization was initially surprising because the cytoplasm of both prokaryotic and eukaryotic cells usually destabilizes globular proteins [67]. Any small differences between intracellular pH and the buffered dilute solutions cannot explain the stability difference because dimer stability changes only by ~0.1 kcal/mol between pH 6.2 and 7.5 [9] and the stability of GB1 is insensitive to physiologically-relevant ionic strength [88]. We suggest that the stabilization in cells arises mostly from repulsive chemical interactions between the client protein and the cellular milieu.

The idea that increased dimer stability in cells arises from hard-core repulsions seems unlikely for two reasons. First, we know from *in vitro* crowding studies that the A34F variant is nearly insensitive to hard-core repulsive effects [60]. Second, if we interpret the results using traditional theories and bear in mind that *E. coli* cells are more crowded than oocytes [69], the expectation is that the dimer would be more stable in *E. coli*, but the opposite is true. This analysis suggests that chemical interactions play a

key role in protein-complex stability in cells, and, specifically, the stabilizing effect in both cell types arises from the repulsion between A34F protein, which has a charge of -4 in cells and the overall net-negative charge on the intracellular proteins [70, 89].

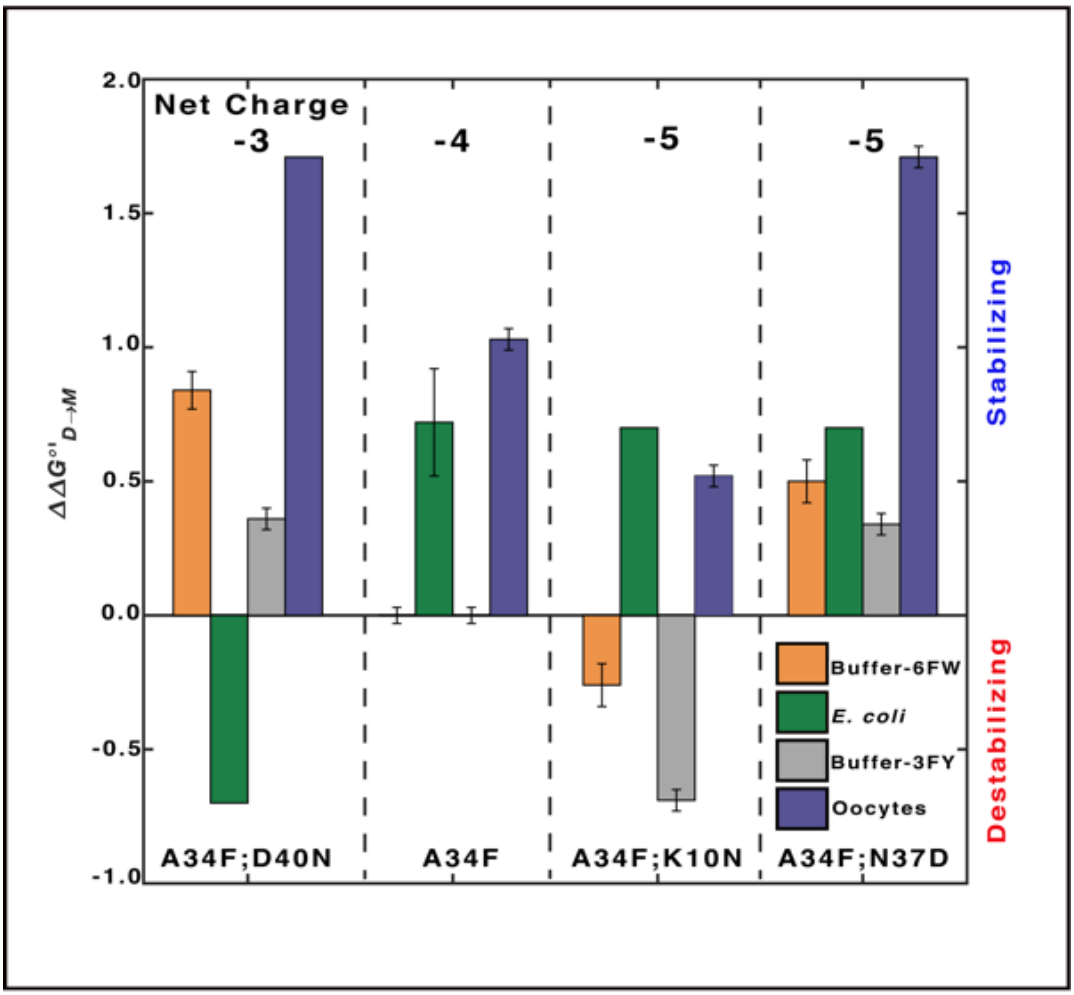

Figure 4.3. Charge and dimer stability (**$\Delta\Delta G^{o'}_{D\rightarrow M} = \Delta G^{o'}_{D\rightarrow M,variant,buffer/cells} - \Delta G^{o'}_{D\rightarrow M,A34F,buffer}$**). Positive values indicate increased stability and *vice versa*. Uncertainties are propagated from the uncertainties in $\Delta G^{o'}{}_{D\rightarrow M}$. The absence of error bars indicates minimum magnitude of $\Delta\Delta G^{o'}_{D\rightarrow M}$for variants exhibiting only dimer or monomer in cells.

To test the role of electrostatic interactions between GB1 and the cellular milieu, we studied variants that alter the surface charge, comparing their stabilities to that of the A34F protein in buffer (i.e., $\Delta\Delta G^{o\prime}_{D\rightarrow M} = \Delta G^{o\prime}_{D\rightarrow M,variant,buffer/cells} - \Delta G^{o\prime}_{D\rightarrow M,A34F,buffer}$, Fig. 4.3). The A34F;D40N variant has less charge, -3, than wild-type A34F, -4. Two of the variants A34F;K10N and A34F;N37D have a greater charge, -5, than wild-type A34F GB1, -4.The estimated net charge, predicted using Protcalc, on each protein at physiologically relevant pH values is shown at the top of Fig. 4.3. The changes are all in loops [61] and are expected to have a small effect on structure. Inspection of ^{15}N-^{1}H HSQC acquired in buffer alone (figs. S4.9 and S4.10) confirm this expectation, showing indication of dimer formation by signature resides such as T17. For certain variants only the monomer or dimer is observed in cells (fig. S4.11 and S4.12). For these proteins, we used the minimum or maximum detectable stability (table S4.1) to estimate the minimum magnitude of $\Delta\Delta G^{o\prime}_{D\rightarrow M}$ Fig. 4.3. Therefore, the actual positive and negative values of $\Delta\Delta G^{o\prime}_{D\rightarrow M}$are larger and smaller than the estimated values, respectively.

First, we consider $\Delta\Delta G^{o\prime}_{D\rightarrow M,variant,buffer}$ in buffer. Both 6FW- and 3FY- labeled proteins (orange and gray bars, respectively) show the same trends, indicating that it is reasonable to compare the differently labeled proteins. Two of the three variants behave as anticipated in buffer based on charge. For A34F;D40N, the decrease in charge-charge repulsions between monomers compared to A34F (-4 to -3), increases dimer stability. For A34F;K10N, the increase in repulsions between monomers (-4 to -5) decreases stability. Contrary to this simple idea, A34F;N37D variant (also (-4 to -5) is more stable than A34F, but the side chain of residue 37 is near the dimer interface [61], so there may be slight perturbations to structure. Both A34F;K10N and A34F;N37D

make hydrogen bonds for residues within the dimer interface, but A34F;D40N has no polar contacts to any other atom.

The behavior of the charge-change variants in both *E. coli* cells and oocytes is consistent with fact that most cellular proteins possess a net negative charge [70, 89]. Specifically, if charge-charge interactions are important, we expect that the more negative the charge on the homodimer, the stronger the intermolecular repulsion in cells and the more stable the complex. This prediction is borne out by the data for both cell types, even though some of the stabilities cannot be quantified because only the dimer or monomer is observed. Furthermore, their increased stability in oocytes occurs despite the fact that the g/L protein concentration in oocytes is only about half that of *E. coli* [69]. These observations show that chemical interactions are a key determinant of protein behavior in cells.

We can also integrate our knowledge about protein- and protein-complex stability in cells by considering the shapes of the products and reactants. Early expectations for protein stability were based on ideas that globular proteins should stabilized under crowded conditions because the native state occupies less space than the unfolded state. The validity of these ideas, however, depends in some sense on treating both states as hard, sphere-like objects, which certainly is not the case for the unfolded state. Unfolding proteins also have exposed sites for attractive interactions such as hydrogen bonding and hydrophobic contacts, which lead to destabilization. For protein-protein interactions such as the one studied here, both the reactants and products can be considered more like hard objects such that attractive and repulsive interactions can be simply rationalized. This is further supported by a recent study showing A34F is barely

affected by hard-core repulsions in dilute buffer. Perhaps the strongest pieces of evidence for the importance of charge-charge interactions is cells are that all the charge-change variants have essentially the same excluded volume, yet show radically different behaviors in cells and the fact that excluded volume arguments only predict increases in stability under crowded conditions, yet the A34F;D40N variant is less stable in cells than it is in buffer. A34F;D40N may be less stable in *E. coli* cells because it is less charged it may experience more attractive chemical interactions.

Protein-protein interactions allow cells to respond to chemical and physical stimuli, but non-specific interactions in the crowded cellular interior inevitably compete with specific interactions and interfere with signal transduction. Protein surface charge has been shown to play a determinant role for protein diffusion in cells due to non-specific interactions within the cytoplasm [90-92]. Our study shows that increasing the negative charge on a protein complex can enhance a specific protein-protein interaction. The data also suggest a reason why post translational modifications such as methylation, acetylation, and protein phosphorylation, which adds negative charge, are used to control signaling. Charge-charge interactions are also important for proper protein function. For instance, a single surface-charge change causes sickle cell disease [93] and myoglobin surface charge correlates with diving ability in mammals [94], and the charge on the loops regions in crystallin are key to eye lens formation [95]. Therefore, investigating how the cellular environment tunes specific protein-protein interactions is crucial to advancing our physical understanding of biology and human health.

APPENDIX 1: SUPPORTING INFORMATION

Plasmids

pET-11a harboring the gene for the T2Q;A34F mutant of GB1 was used as the basis of the project. The T2Q change prevents N-terminal degradation, [96] and the A34F change causes dimer formation [61]. The GB1 proteins discussed here carry these two changes. Agilent Quick-Change kits, A34F;D40N, or gene synthesis (Gene Universal), A34F;K10N and A34F;N37D was used to construct other mutants.

6-Fluorotryptophan (6-FW) Labeled Protein

Protein was expressed and purified as described [82]. Briefly, a 1-L culture of *E. coli* strain BL21 DE3 harboring a GB1 construct was grown in ampicillin-containing (100 µg/L final concentration), ^{15}N-enriched M9 media (6.78 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/L $^{15}NH_4Cl$, 4 g/L D-glucose, 2 mM $MgSO_4$, 1 mM $CaCl_2$, 1 mM ampicillin) with shaking at 37 °C at 225 rpm (New Brunswick Scientific, Innova I26). When the cells reached an optical density at 600 nm (OD_{600}) of 0.6 – 0.8, 1 g *N*-(phosphonomethyl)glycine, 60 mg phenylalanine, 80 mg tyrosine were added. Next, 60 mg of 6-fluoroindole (6-FI; Sigma-Aldrich) was dissolved in 250 µL of dimethyl sulfoxide (DMSO) and added. After shaking for an additional 45 min, protein expression was induced with isopropyl β-$_D$-1-thiogalactopyranoside (IPTG) at a final concentration of 1 mM.

After 2 h, the cells were pelleted at 4000*g* for 30 min. The supernatant was discarded, and the cells were lysed by sonication (Fischer Scientific Sonic Dismembrator model 500, 15% amplitude, 10 min, 0.50 s on, 0.50 s off) in 30 mL of 20 mM tris (pH 7.5) containing protease inhibitor (Roche, cOmplete, EDTA-free inhibitor cocktail). The lysate was centrifuged at 15000*g* for 45 min to remove cell debris. The supernatant was filtered (0.45 µm) and loaded on a 16 mm x 25 mm Q Sepharose anion exchange column attached to a GE ÄKTA FPLC. Protein was eluted over a 0-50% linear gradient of 20 mM tris (pH 7.5) to 20 mM tris, 2 M NaCl (pH 7.5) at 277 K. Fractions were assessed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) with Coomassie blue staining. Fractions containing GB1 were concentrated

using a 3000 Da Amicon centrifugal concentrator. The concentrated sample was loaded on a 16 mm x 600 mm GE Superdex-75 size exclusion column and eluted with 1X M9 salts (3 mM Na_2HPO_4, 2 mM KH_2PO_4, 9 mM NaCl, pH 7.5 at 277 K).

Fractions were assessed with SDS-PAGE using Coomassie staining. Purified fractions containing GB1 were concentrated using a 3000 Da cut off Amicon centrifugal concentrator, and buffer exchanged into sterilized deionized H_2O (18-MΩ). Protein concentration was quantified using a NanoDrop One Spectrophotometer (ThermoFisher) and an extinction coefficient of 9970 L M^{-1} cm^{-1} [97]. Purified protein was split into 500 µM aliquots, lyophilized for 12-16 hours, and stored at -20°C. Protein was resuspended in 20 mM sodium phosphate buffer (pH 7.5) containing 10% D_2O for NMR experiments.

3-Fluorotyrosine (3-FY) Labeled Protein

^{15}N-enriched, 3-fluorotyrosine labeled protein was prepared as described [86, 98]. Briefly, a single colony of *E. coli* strain BL21 DE3 containing the appropriate GB1-harboring plasmid was inoculated in 10 mL of LB media (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 1 mM ampicillin). The culture was shaken overnight at 37 ºC and transferred to 100 mL of tryptone-yeast media (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, 1 mM NaOH, 1 mM ampicillin). After 2 h of shaking at 37 ºC, the cells were centrifuged and the pellet resuspended in 1 L of ^{15}N-enriched M9 media and shaken at 37 °C. When the cells reached an OD_{600} of 0.4, 0.5 g of glyphosate, 70 mg of 3-fluorotyrosine, 60 mg of tryptophan, and 60 mg of phenylalanine were added. The culture was then grown to an OD600 of 0.6, and protein expression was induced with 1 mM isopropyl IPTG. After 2 hr the pellet was collected by centrifugation at 6000 rpm (JA-10 rotor, Beckman coulter) for 10 min and stored at -20 ºC.

The pellet was resuspended in 20 mM tris, pH 7.5. The lysate was obtained from sonication (Scientz-IID JY92-IIN, 40% amplitude, 30 min, 3 s on, 6 s off) was centrifuged for 30 min at 20000 rpm (JA-25.5 rotor, Beckman coulter). The supernatant was loaded on a 16 mm x 100 mm anion exchange column (GE HiPrep DEAE FF 16/10) attached to a GE ÄKTA FPLC. The protein was eluted with a gradient from 0 to 1

M NaCl. The fractions were assessed by SDS-PAGE with Coomassie staining. GB1-containing fractions were concentrated using a 3000-molecular-weight-cutoff Amicon centrifuge filter and loaded on a 26 mm x 600 mm GE Superdex-100 size exclusion column equilibrated with 20 mM tris, 250 mM NaCl, pH 7.5 at 277 K. GB1 containing fractions were exchanged into triply-distilled H_2O (1.5 MΩ cm) with a desalting column (GE HiPrep 26/10, Sephadex G-25F). Purity was assessed with SDS-PAGE and Coomassie staining. Pure GB1 containing fractions were concentrated and buffer exchanged into sterilized deionized H_2O. Protein concentration was quantified at 280 nm as described above. The purified protein was lyophilized (Alpha 1-4 LD plus, Martin Christ) and stored at -20 ºC. Dimer stability in buffer was assessed as described by Guseman and Pielak [82].

E. coli In-cell NMR

Samples were prepared as described [83]. Briefly, a plasmid containing the gene for a GB1 variant was transformed into Tuner (DE3) cells (Novagen) by heat shock. A single colony was used to inoculate 5 mL of LB media supplemented with 100 µg/L ampicillin. After the culture was shaken at 37 °C at 225 rpm for 6–8 h, 500 µL was used to inoculate 100 mL of M9 media. The 100 mL culture was incubated with shaking at 37 °C overnight and added to 100 mL of fresh supplemented M9 media. The culture was incubated at 37 °C. When the OD_{600} reached 0.6 – 0.8, 12 mg of 6-fluoroindole dissolved in 250 µL DMSO was added. After shaking for an additional 45 min, IPTG was added to induce protein expression. After 45 min, 50 µg/L chloramphenicol was added to halt protein expression.

Three 1 mL aliquots were taken from the in-cell culture for liquid chromatography mass spectrometry to determine protein concentration as described [63]. The cells were pelleted at 2000*g* for 20 min and washed three times with an in-cell NMR buffer (200 mM HEPES and 100 mM bis-tris propane dissolved in 10% D_2O, pH 7.6). The pellet was resuspended in 250 µL of in-cell NMR buffer. To check for leakage, after each in-cell spectrum was obtained, the sample was gently pelleted, and a spectrum of the 2-fold diluted supernatant was acquired. No leakage was observed. Example supernatant samples were similar to those reported by Chu et al. [83].

E. coli In-cell NMR HSQCs

Samples were prepared as described [83]. Briefly, a plasmid containing the gene for a GB1 variant was transformed into Tuner (DE3) cells (Novagen) by heat shock. A single colony was used to inoculate 50 mL of LB media supplemented with 100 µg/L ampicillin. After the culture was shaken at 37 °C at 225 rpm for 6–8 h, the cells were pelleted at 2000*g* (Sorvall ST 16 Centrifuge) for 10 min. The pellet was then resuspended with 200 mL of supplemented M9 media. The culture was incubated at 37 °C. When the OD_{600} reached 0.8, 12 mg of 6-fluoroindole dissolved in 250 µL DMSO was added. Then the cells were shaken until an OD_{600} of 1.2 was reached, 1 mM of IPTG was added to induce protein expression. After 3 h, 50 µg/L chloramphenicol was added to halt protein expression.

Three 1 mL aliquots were taken from the in-cell culture for liquid chromatography mass spectrometry to determine protein concentration as described [63]. The cells were pelleted at 2000*g* for 10 min. The pellet was resuspended in 250 µL of in-cell NMR buffer (1 x M9 in 10% D_2O, pH 7.5). To check for leakage, after each in-cell spectrum was obtained, the sample was gently pelleted, and a spectrum of the 2-fold diluted supernatant was acquired. No leakage was observed. Example supernatant samples were similar to those reported by Chu et al. [83].

Preparing Xenopus laevis oocytes for NMR

Oocytes were prepared as described [99]. Briefly, ovary lobes were surgically removed to ND96-Ca^{2+} buffer (96 mM NaCl, 2 mM KCl, 1 mM $MgCl_2$, 5 mM HEPES, pH 7.4). After washing, the ovaries were digested with collagenase (2 mg/mL final concentration). Oocytes were then washed in ND96 (96 mM NaCl, 2 mM KCl, l mM $MgCl_2$, 5 mM HEPES, pH 7.4) and then ND96 containing 1.8 mM $CaCl_2$. Stage-Ⅵ oocytes were selected for microinjection. The protein solution (about 40 nL contain 30% v/v P1860-protease inhibitor cocktail (Sigma-Aldrich)) was injected into each oocyte *via* an IM-300 microinjector (Narishige Co. Ltd., Tokyo, Japan). About 100 oocytes in ND96 buffer containing 10% D_2O were transferred to a Shigemi micro-NMR tube.

NMR of 6FW-labeled Proteins

^{19}F NMR experiments were performed at 298 K on a 500 MHz Bruker Avance III HD spectrometer equipped with a QCI cryogenic probe. Data were analyzed using TopSpin 3.6.1. Spectra comprised 32768 points, 512 scans, an acquisition time of 0.9 s, and a sweep width of 20 ppm. The ^{15}N-^{1}H HSQC spectra comprised of 2048 points ^{15}N and 128 points ^{1}H, 64 scans, an acquisition time of 0.12 s for ^{15}N and 0.02 s for ^{1}H and a sweep width of 44 ppm for ^{15}N and 13 ppm for ^{1}H. Data were processed with nmrPipe and NMRViewJ. The spectra were acquired in 200 mM Hepes, 100 mM bis-tris propane, 150 μg/mL ampicillin and 50 μg/mL chloramphenicol in 10% D_2O (pH 7.6). The GB1 concentration was varied from 20 μM to 2 mM.

HSQCs of 6FW-labeled Proteins in Living Cells

^{15}N-^{1}H HSQC spectra experiments were performed at 298 K on a 600 MHz Bruker Avance III HD spectrometer. Data were analyzed using TopSpin 3.6.2. Spectra comprised of 2048 points ^{15}N and 128 points ^{1}H, 32 scans, an acquisition time of 0.12 s for ^{15}N and 0.02 s for ^{1}H and a sweep width of 44 ppm for ^{15}N and 13 ppm for ^{1}H. Data were processed with nmrPipe and NMRViewJ. The spectra were acquired in 1 X M9 and 50 μg/mL chloramphenicol in 10% D_2O (pH 7.5). The GB1 concentration was 2 mM.

NMR of 3FY-labelled proteins

Experiments were performed at 288 K on a 600 MHz spectrometer (Bruker) equipped with H/F/(C, N) triple-resonance cryogenic probe. ^{19}F spectra were acquired with a sweep width 11.31 kHz, a relaxation delay 2.5 s and an acquisition time of 1.45 s. The number of pulses was 1024. The spectral width of 1D ^{1}H-^{15}N HSQC spectra was 16 ppm for ^{1}H, a relaxation delay 1.5 s and an acquisition time of 0.11 s. The number of pulses was 512. For ^{1}H-^{15}N HSQC, the spectral width was 16 ppm for ^{1}H and 40 ppm for ^{15}N, a relaxation delay of 1.5 s and an acquisition time of 0.11 s. The number of pulses was 16. The spectra were acquired in 20 mM sodium phosphate in 10% D_2O at pH 7.4. The GB1 concentration was varied from 50 μM to 1.5 mM.

Data analysis

Data were analyzed using TopSpin 3.6.1 or 3.2 with line-broadening of ^{19}F and 1D ^{1}H-^{15}N HSQC spectra of 4.0 Hz and 0.3 Hz, respectively. Peak fitting of ^{19}F spectral and the area of dimer and monomer peak was accomplished with Topspin or OriginPro 8.5.1. Relative populations of dimer and monomer were obtained via integrations. The fraction of dimer (F_d) was calculated by dividing the integral of the dimer peak by the sum of the integrals of the monomer and dimer peaks. Data were fit to Eq. 1 using MATLAB (R2020A) or Origin.

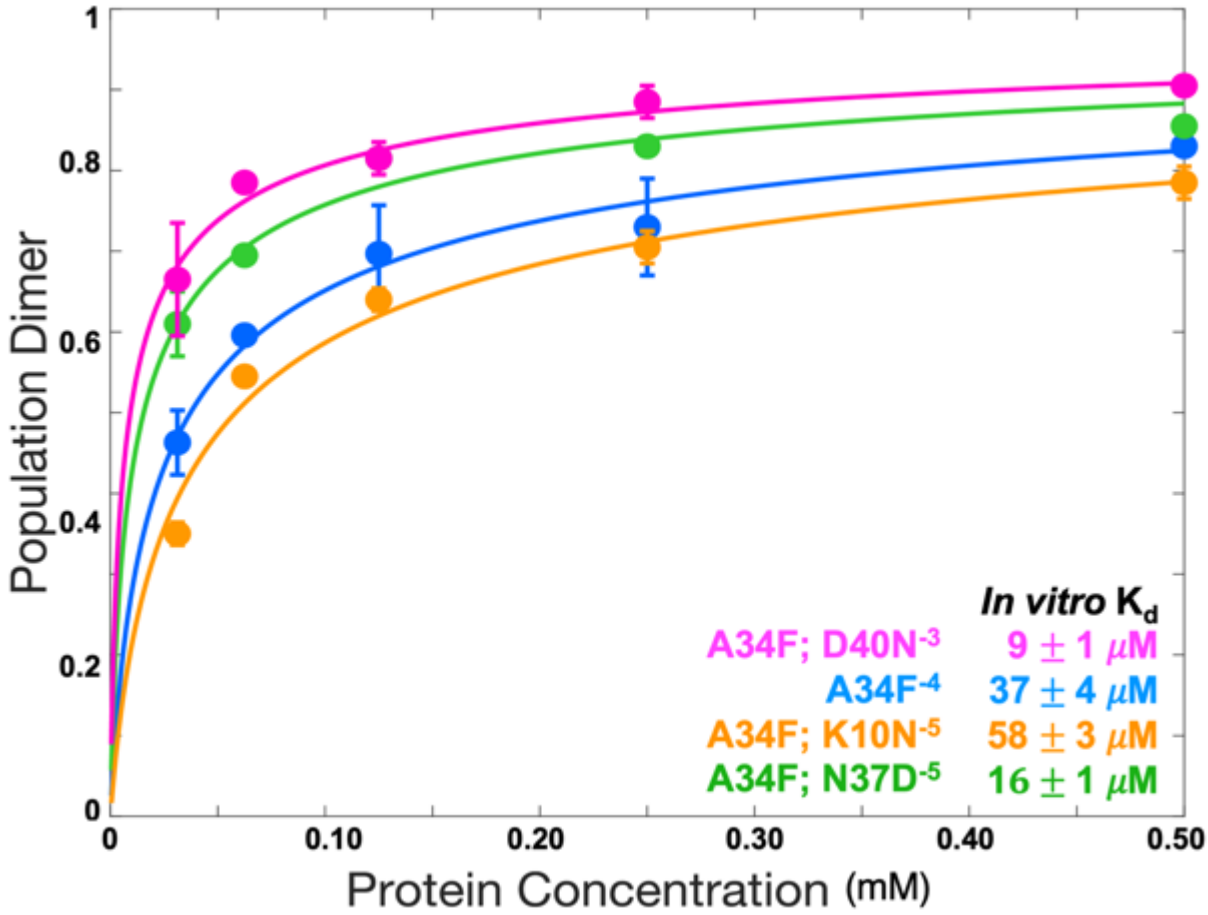

Figure S4.1. Quantifying dissociation of 6-fluorotryptophan labeled A34F^{-4}, A34F; N37D^{-5}, A34F; D40N^{-3}, and A34F; K10N^{-5} GB1 in buffer. Uncertainties are the standard deviation of the mean from triplicate experiments. Superscripts denote net charge at pH 7.5.

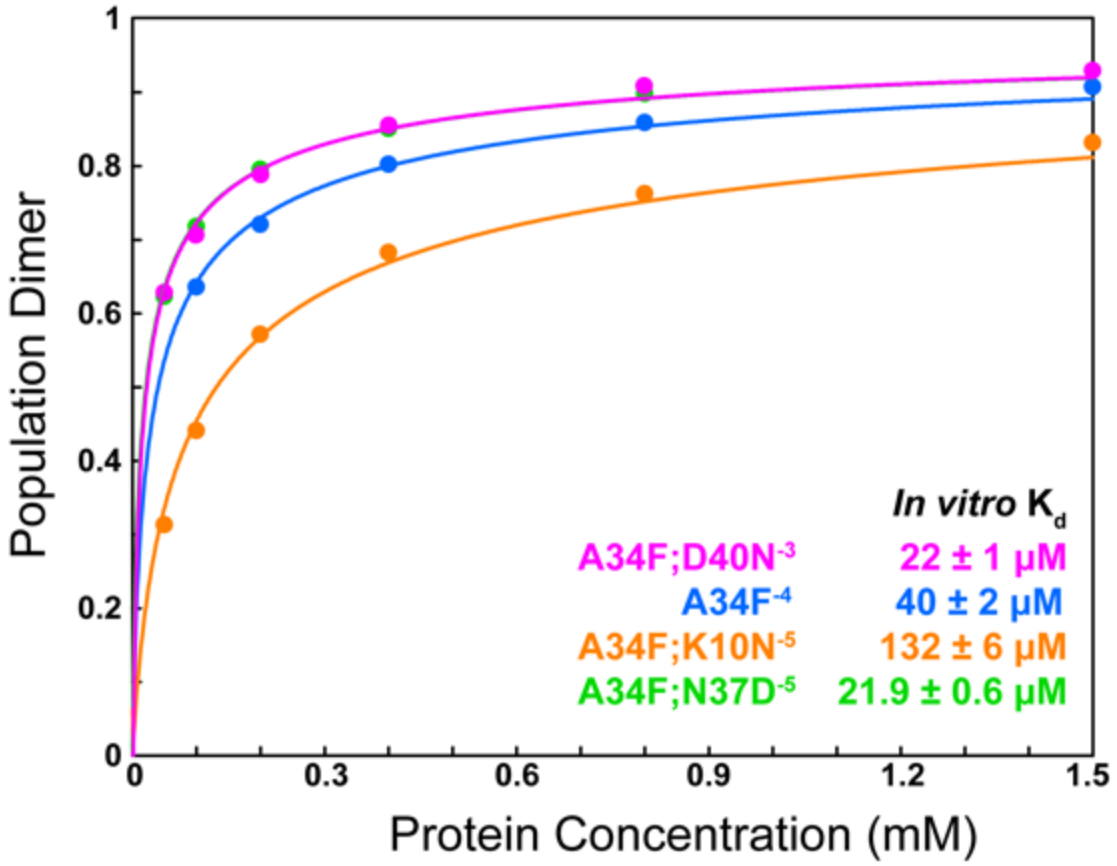

Figure S4.2. Quantifying dissociation of 3-fluorotyrosine labeled A34F^{-4}, A34F; N37D^{-5}, A34F; D40N^{-3}, and A34F; K10N^{-5} GB1 in buffer. Superscripts denote net charge at pH 7.5. Uncertainties are derived from least squares fitting.

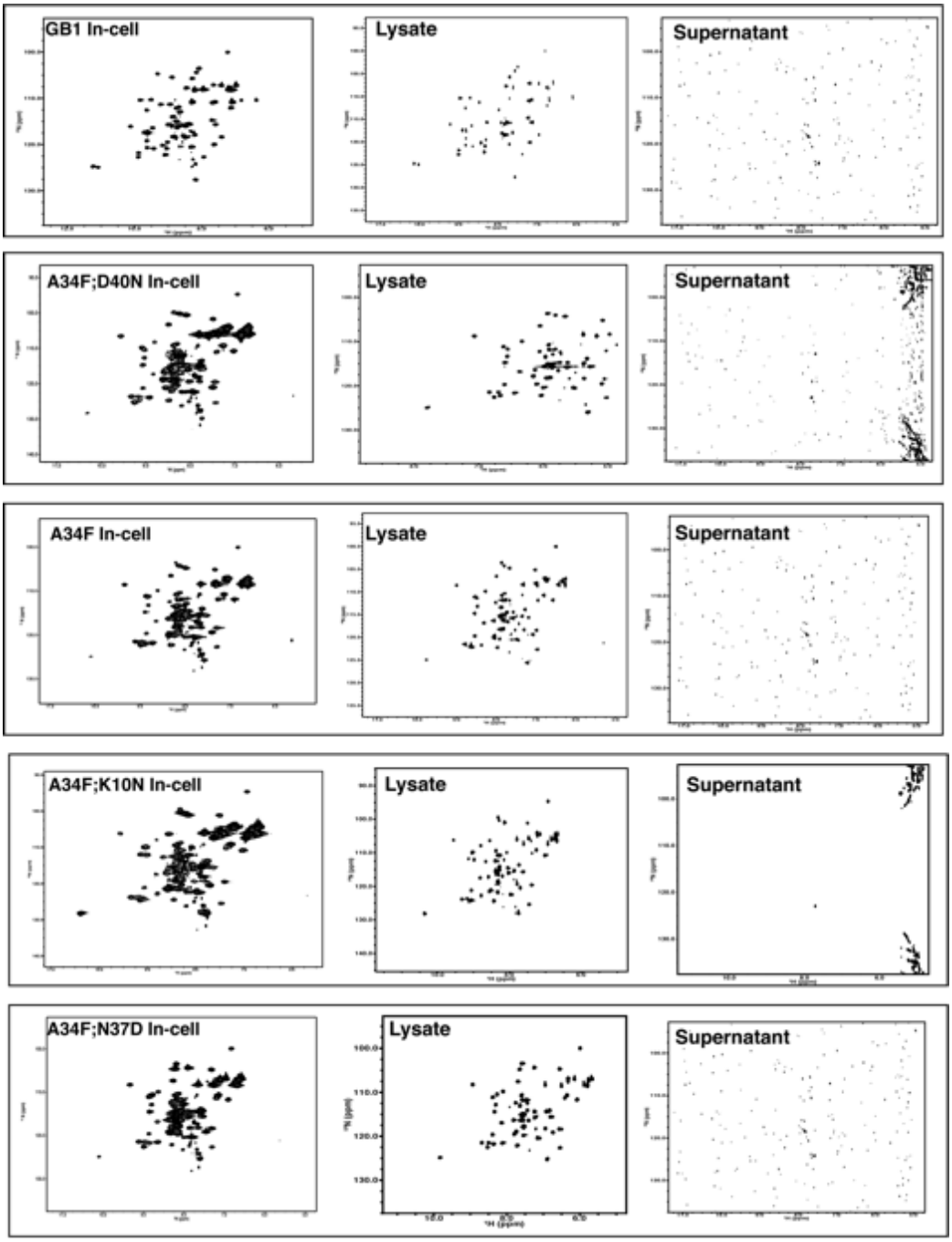

Figure S4.3. ^{15}N-^{1}H HSQC of ^{19}F-Trp lad GB1^{-4}, A34F; D40N^{-3}, A34F^{-4}, A34F; K10N^{-5}, and A34F; N37D^{-5} in living *E. coli* cells at 298 K pH 7.5. Lysate and supernatant controls verify the protein of interest is being measured inside cells. Supernatant spectra are zoomed in to show there is only noise and no protein leakage.

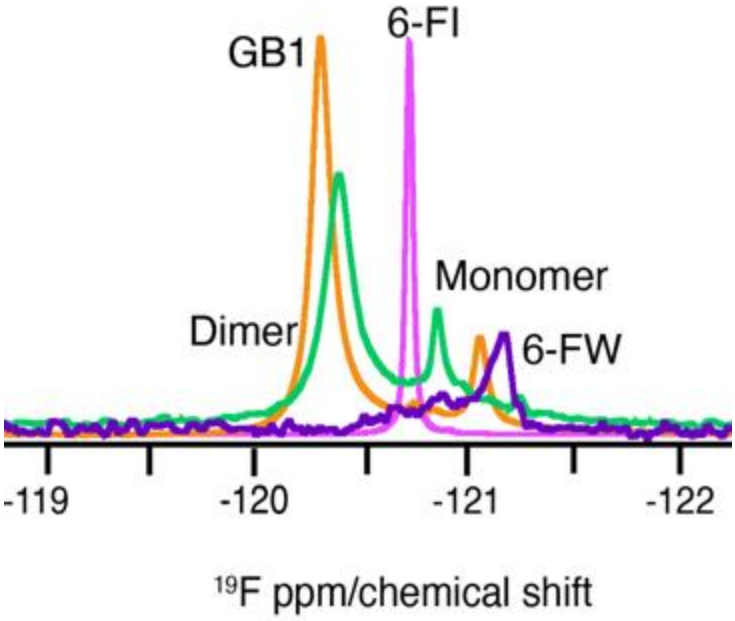

Figure S4.4. ^{19}F spectra of GB1 (orange), GB1; A34F induced in living *E. coli* cells (green) verify that there is no overlap of monomer and 6-FW inside cells. GB1; A34F without inducer confirms the second resonance in the GB1 spectrum is 6-FW (purple). Control spectrum of 6-FI in buffer (magenta) confirms there is no overlap with the monomer or dimer in cells.

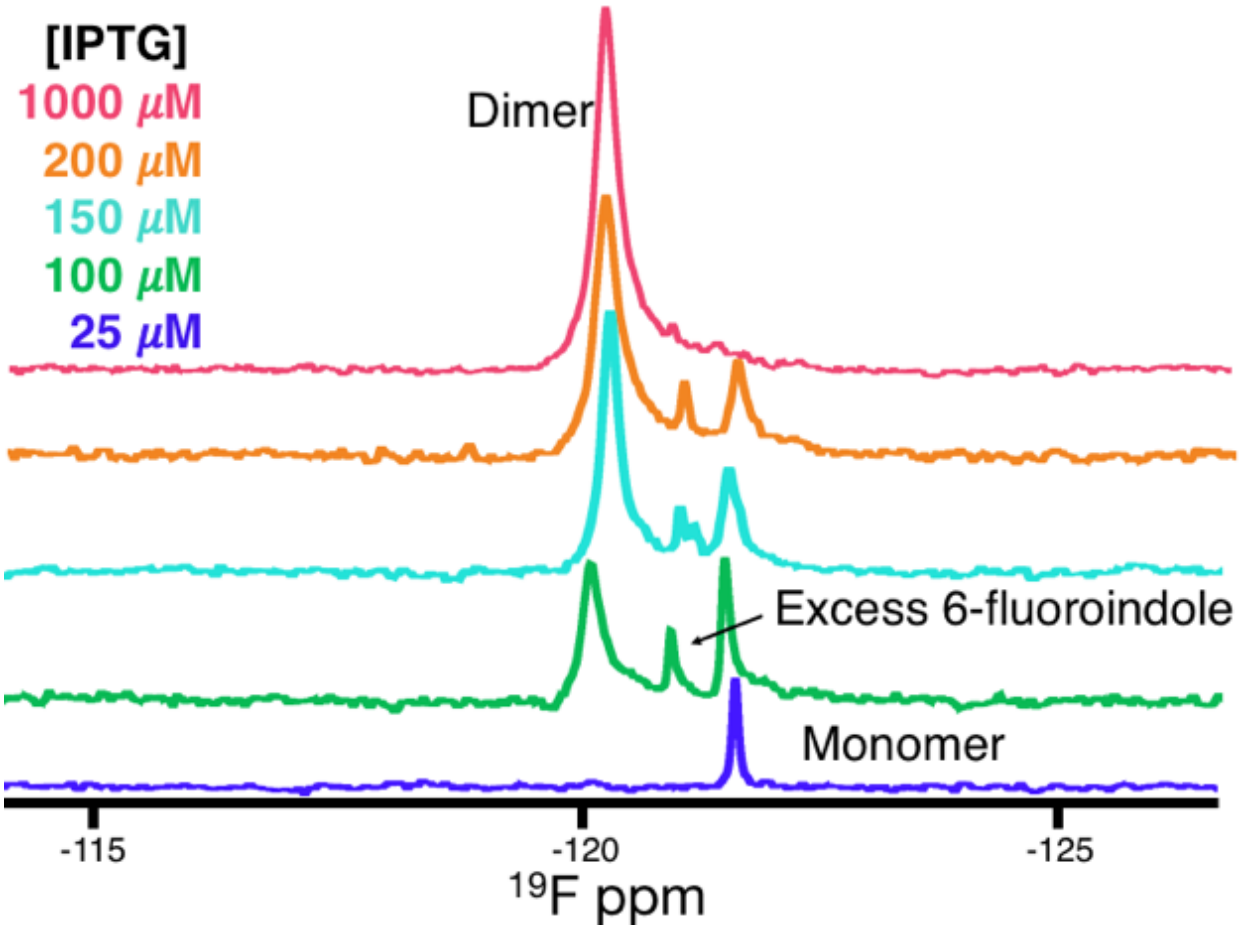

Figure S4.5. ^{19}F NMR Spectra of 6-fluoroindole-labeled A34F GB1 in *E. coli* Tuner cells as a function of inducer (IPTG) concentration.

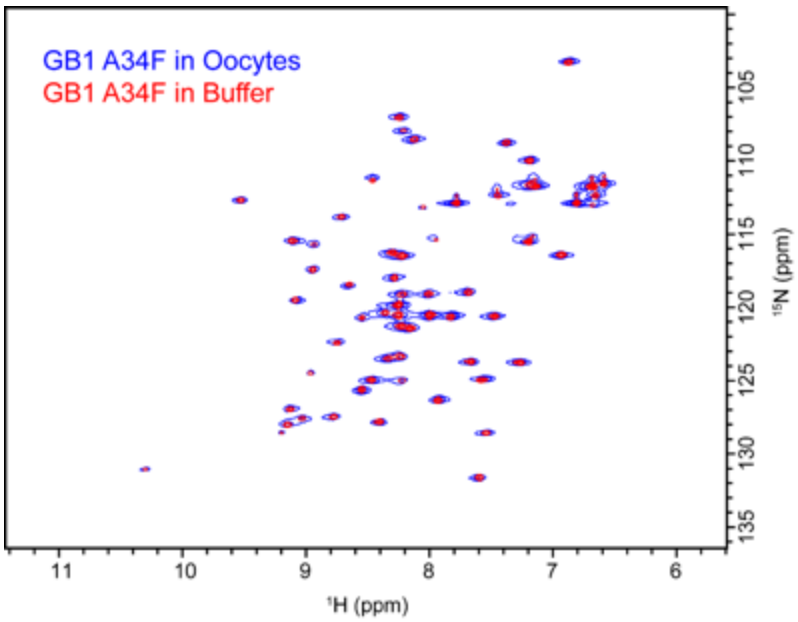

Figure S4.6. ^{1}H-^{15}N HSQC spectra in oocytes (blue) and in buffer (red, 20 mM phosphate buffer, pH 7.4) of ^{15}N-enriched, 3FY-labeled A34F GB1.

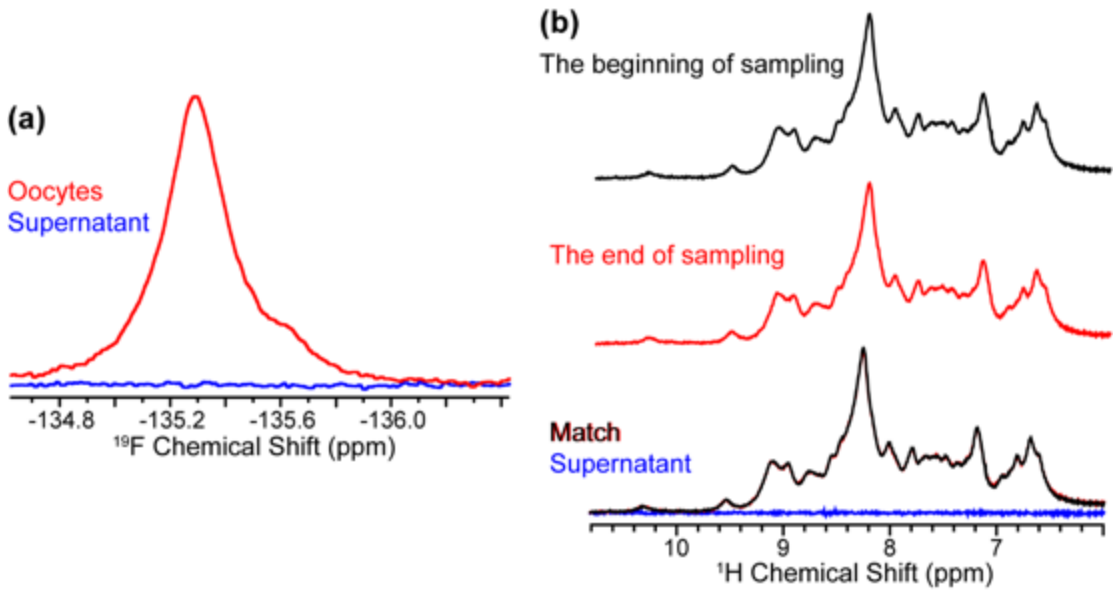

Figure S4.7. A34F GB1 is stable in oocytes for the duration of the NMR experiment, and there was no leakage. (a) ^{19}F NMR spectra of A34F GB1 in oocytes (red) and supernatant (blue). (b) 1D ^{1}H-^{15}N HSQC spectra of A34F GB1 in oocytes before (black) and after (red) acquisition of the^{19}F NMR spectrum.

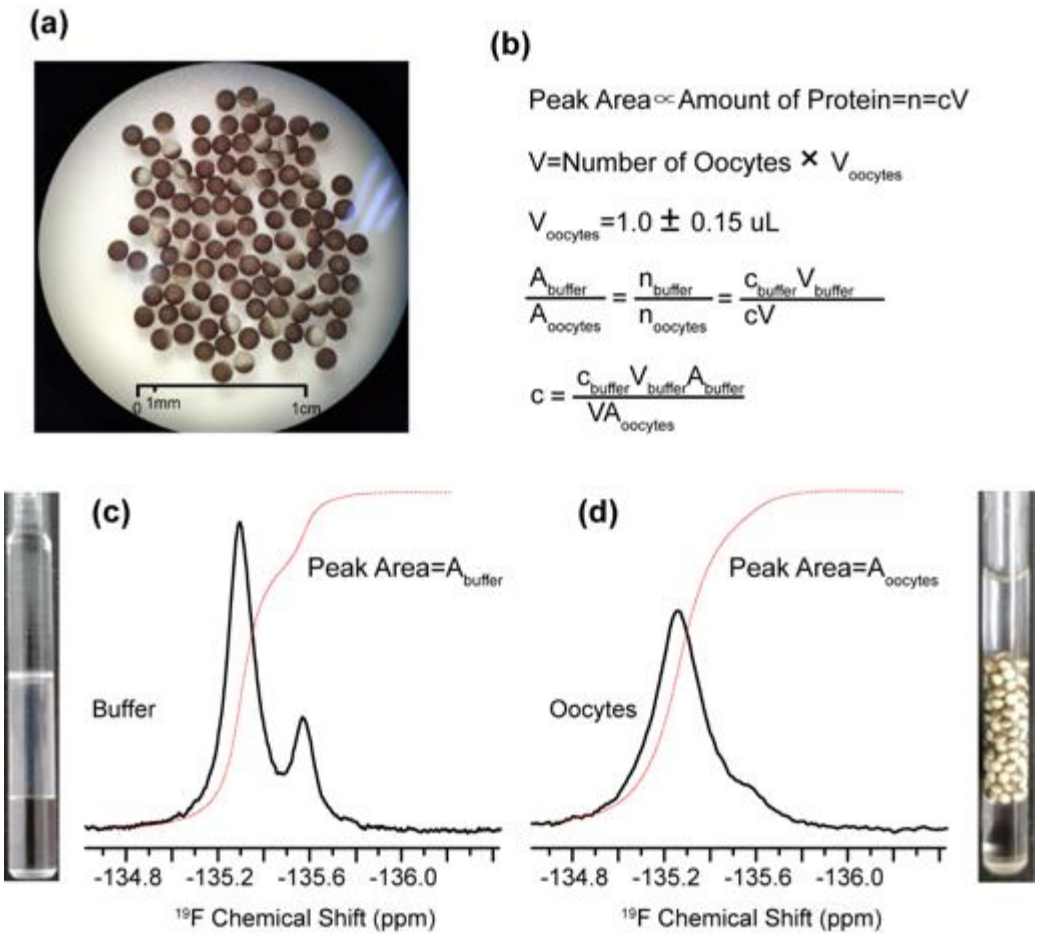

Figure S4.8. Concentration of A34F GB1 in oocytes. (a) Oocytes for in-cell NMR experiments. (b) Equation to determine intraoocyte A34F concentration [A_{buffer} ($A_{oocytes}$), area of ^{19}F NMR resonance in buffer; n_{buffer} ($n_{oocytes}$), moles of total protein; c_{buffer} ($c_{oocytes}$), concentration of protein in buffer (oocytes); V_{buffer} ($V_{oocytes}$), volume of buffer]. (c and d) Buffer and oocytes containing A34F GB1 in Shigemi micro-NMR tubes and their integrated ^{19}F NMR spectra.

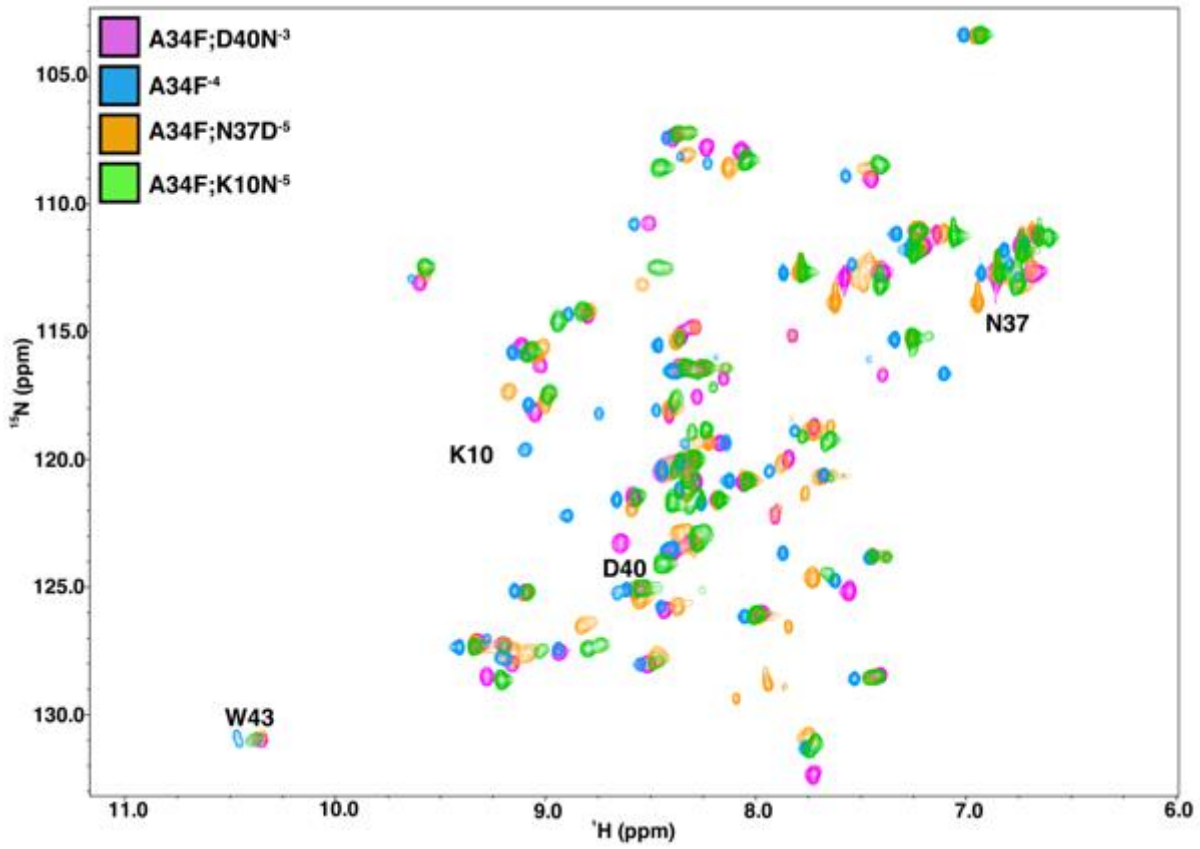

Figure S4.9. $^{15}N^{-1}H$ HSQC spectra of 6FI-labeled proteins in buffer alone. Superscripts denote net charge at neutral pH.

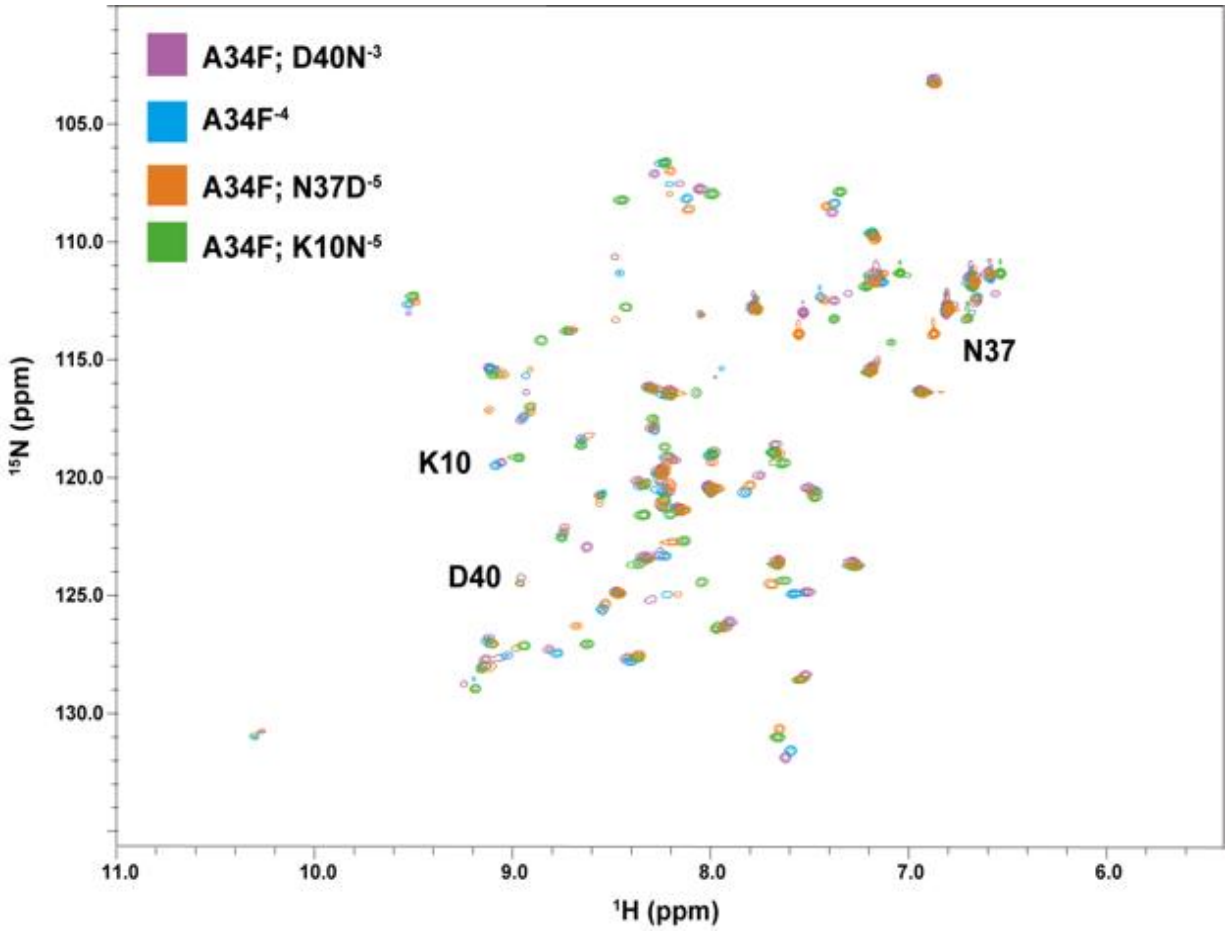

Figure S4.10. $^{15}N^{-1}H$ HSQC spectra of 3FY-labeled proteins in buffer alone. Superscripts denote net charge at neutral pH.

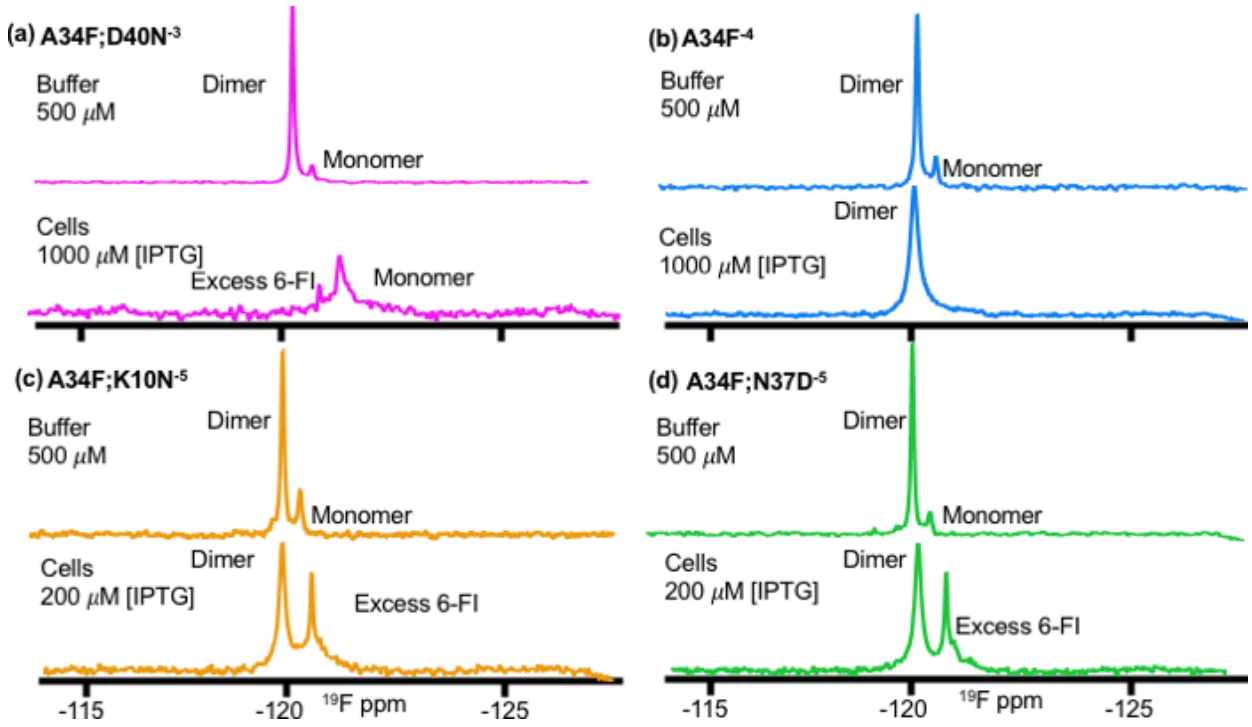

Figure S4.11. ^{19}F NMR spectra in buffer and *E. coli* of 6-fluorotryptophan variants that show only monomer or dimer in cells: A34F; D40N^{-3} (a), A34F^{-4} (b), A34F; K10N^{-5}, (c) A34F; N37D^{-5}. Superscripts denote net charge at pH 7.5.

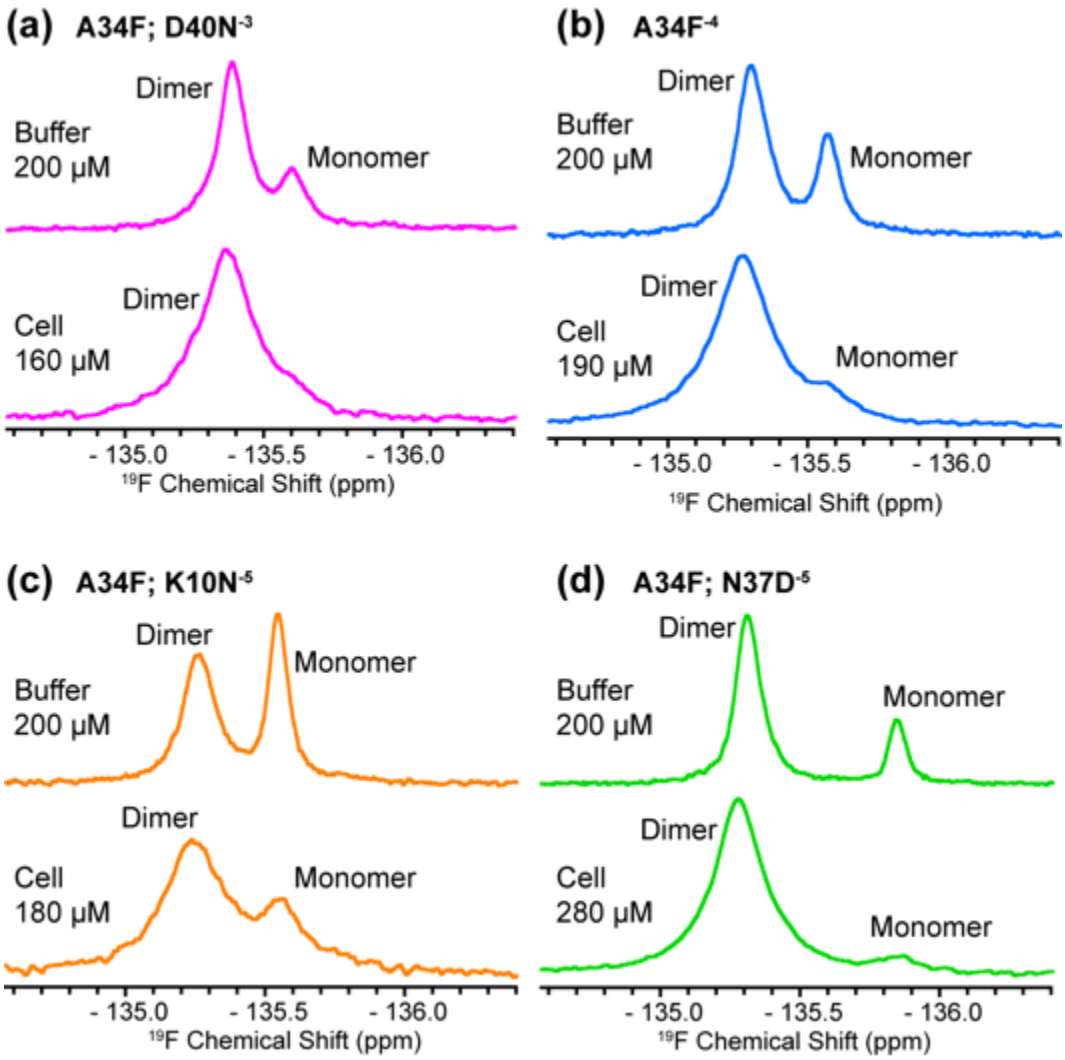

Figure S4.12. ^{19}F spectra of 3-fluorotyrosine labeled A34F;D40N^{-3} (a) , A34F^{-4} (b) , A34F;K10N^{-5} (c) and A34F;N37D^{-5} (d) in buffer (red) and oocytes (blue). Superscripts denote net charge at pH 7.5.

Table S4.1. Equilibrium dissociation parameters at 298 K for 6-fluorotryptophan- and 288 K for 3-fluorotyrosine- labeled proteins.

Condition	$K_{D\rightarrow M}(\mu M)$	$\Delta G^{\circ\prime}_{D\rightarrow M}$ (kcal/mol)	$\Delta\Delta G^{\circ\prime}_{D\rightarrow M}$[a] (kcal/mol)
A34F; D40N[-3,b], pH 7.6, 6-fluorotryptophan labeled			
20 mM NaP[c]	9 ± 1[d]	6.88 ± 0.04	0.84 ± 0.07[d]
E. coli	only monomer	< 6	< 0.7[e]
pH 7.4, 3-fluorotyrosine labeled			
20 mM NaP	22 ± 1	6.14 ± 0.03	0.34 ± 0.04
Oocytes	only dimer	> 7[f]	> 2[f]
A34F[-4], pH 7.6, 6-fluorotryptophan labeled			
20 mM NaP	37 ± 4	6.04 ± 0.06	0
E. coli	11 ± 4	6.8 ± 0.2	0.7 ± 0.2
pH 7.4, 3-fluorotyrosine labeled			
20 mM NaP	40 ± 2	5.80 ± 0.03	0
Oocytes	6.5 ± 0.7	6.83 ± 0.06	1.03 ± 0.07
A34F; K10N[-5], pH 7.6, 6-fluorotryptophan labeled			
20 mM NaP	58 ± 3	5.78 ± 0.06	-0.26 ± 0.08
E. coli	only dimer	> 7[e]	> 0.7[e]
pH 7.4, 3-fluorotyrosine labeled			
20 mM NaP	132 ± 6	5.11 ± 0.03	-0.69 ± 0.04
100 g/L Lysate	48 ± 3	5.69 ± 0.03	-0.11 ± 0.04
Oocytes	16 ± 2	6.31 ± 0.05	0.51 ± 0.06
A34F; N37D[-5], pH 7.6, 6-fluorotryptophan labeled			
20 mM NaP	16 ± 1	6.54 ± 0.06	0.50 ± 0.08
E. coli	only dimer	> 7[e]	> 0.7[e]
pH 7.4, 3-fluorotyrosine labeled			
20 mM NaP	21.9 ± 0.6	6.14 ± 0.02	0.34 ± 0.04
Oocytes	1.7 ± 0.1	7.60 ± 0.03	1.80 ± 0.04

Footnotes

[a] $\Delta\Delta G^{\circ\prime}_{D\rightarrow M} = \Delta G^{\circ\prime}_{D\rightarrow M,var} - \Delta G^{\circ\prime}_{D\rightarrow M,A34F,buffer}$

[b] Superscripts denote net charge at neutral pH

[c] NaP, sodium phosphate buffer

[d] Uncertainties are the standard deviation of the mean from triplicate measurements

[e] Stabilities greater than or less than the detection limit

[f] Not determined because monomer and dimer have similar shifts in oocytes, but lysate data suggest the dimer is more stable in oocytes.

REFERENCES

[1] Truong, K. I., Mitshiko. (2001) The use of FRET imaging microscopy to detect protein-protein interactions and protein conformational changes in vivo, *Curr Opin Struct Biol 11*, 573-578.

[2] Sahni, N., Yi, S., Taipale, M., Fuxman Bass, Juan I., Coulombe-Huntington, J., Yang, F., Peng, J., Weile, J., Karras, Georgios I., Wang, Y., Kovács, István A., Kamburov, A., Krykbaeva, I., Lam, Mandy H., Tucker, G., Khurana, V., Sharma, A., Liu, Y.-Y., Yachie, N., Zhong, Q., Shen, Y., Palagi, A., San-Miguel, A., Fan, C., Balcha, D., Dricot, A., Jordan, Daniel M., Walsh, Jennifer M., Shah, Akash A., Yang, X., Stoyanova, Ani K., Leighton, A., Calderwood, Michael A., Jacob, Y., Cusick, Michael E., Salehi-Ashtiani, K., Whitesell, Luke J., Sunyaev, S., Berger, B., Barabási, A.-L., Charloteaux, B., Hill, David E., Hao, T., Roth, Frederick P., Xia, Y., Walhout, Albertha J. M., Lindquist, S., and Vidal, M. (2015) Widespread Macromolecular Interaction Perturbations in Human Genetic Disorders, *Cell 161*, 647-660.

[3] Ellis, R. J. (2001) Macromolecular crowding: obvious but underappreciated, *Trends Biochem. Sci. 26*, 597-604.

[4] Zimmerman, S. B., and Trach, S. O. (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*, *Journal of Molecular Biology 222*, 599-620.

[5] Stadmiller, S. S., Gorensek-Benitez, A. H., Guseman, A. J., and Pielak, G. J. (2017) Osmotic Shock Induced Protein Destabilization in Living Cells and Its Reversal by Glycine Betaine, *J Mol Biol 429*, 1155-1161.

[6] Guseman, A. J., and Pielak, G. J. (2020) Chapter 12 Protein Stability and Weak Intracellular Interactions, In *In-cell NMR Spectroscopy: From Molecular Sciences to Cell Biology*, pp 188-206, The Royal Society of Chemistry.

[7] Minton, A. P. (1981) Excluded volume as a determinant of macromolecular structure and reactivity., *Biopolymers 20*, 2093-2120.

[8] Sarkar, M., Li, C., and Pielak, G. J. (2013) Soft interactions and crowding, *Biophysical Reviews 5*, 187-194.

[9] Guseman, A. J., Speer, S. L., Perez Goncalves, G. M., and Pielak, G. J. (2018) Surface Charge Modulates Protein-Protein Interactions in Physiologically Relevant Environments, *Biochemistry 57*, 1681-1684.

[10] Cohen, R. D., and Pielak, G. J. (2017) A cell is more than the sum of its (dilute) parts: a brief history of quinary structure, *Protein Sci. 26*, 403-413.

[11] Phillip, Y., Kiss, V., and Schreiber, G. (2012) Protein-binding dynamics imaged in a living cell, *Proc. Natl. Acad. Sci. U. S. A. 109*, 1461-1466.

[12] Jiao, M., Li, H. T., Chen, J., Minton, A. P., and Liang, Y. (2010) Attractive protein-polymer interactions markedly alter the effect of macromolecular crowding on protein association equilibria, *Biophys J 99*, 914-923.

[13] Guseman, A. J., and Pielak, G. J. (2017) Cosolute and Crowding Effects on a Side-By-Side Protein Dimer, *Biochemistry 56*, 971-976.

[14] Minton, A. P. (1993) Macromolecular Crowding and Molecular Recognition, *J. Mol. Recognit. 6*, 211-214.

[15] Hoffman L, W. X., Sanabria H, Cheung Margaret S, Putkey John A, Waxham MN. (2015) Relative cosolute size influences the kinetics of protein-protein interactions, *Biophys J 109*, 510-520.

[16] Phillip, Y., Sherman, E., Haran, G., and Schreiber, G. (2009) Common Crowding Agents Have Only a Small Effect on Protein-Protein Interactions, *Biophys J 97*, 875-885.

[17] Linderstrøm-Lang, K. U. (1952) *Proteins and enzymes, In Lane medical lectures*, Stanford University Press, Stanford California.

[18] Cohen, R. P., G.J. (2016) Electrostatic contributions to protein quinary structure, *J Am Chem Soc 40*, 1319-13142.

[19] Davis, C. M. G., M.; Sukenik, S. . (2018) How does solvation in the cell affect protein folding and binding?, *Curr Opin Struct Biol 48*, 23-29.

[20] Rickard, M. Z., Y.; Gruebele, M.; Pogorelov, T.V. (2019) In-cell protein-protein contacts: transient interactions in the crowd, *J Phys Chem Lett 10*, 5667-5673.

[21] Ferguson MWJ, J. T. (1982) Temperature of egg incubation determines sex in Alligator mississippiensis, *Nature 296*, 850-853.

[22] Maeder CI, H. M., Kinkhabwala A, Mayr R, Bastiaens PIH, Knop M. (2007) Spatial regulation of fus3 map kinase activity through a reaction-diffusion mechanism in yeast pheromone signalling, *Nat Cell Biol 9*, 1319-1326.

[23] Sudhaharan T, L. P., Foo YH, Bu W, Lim KB, Wohland T, Ahmed S. (2009) Determination of in vivo dissociation constant, kd, of cdc42-effector complexes in live mammalian cells using single wavelength fluorescence cross-correlation spectroscopy, *J Biol Chem 284*, 13602-12609.

[24] Shi X, F. Y., Sudhaharan T, Chong S-W, Korzh V, Ahmed S, Wohland T. (2009) Determination of dissociation constants in living zebrafish embryos with single wavelength fluorescence cross-correlation spectroscopy, *Biophys J 97*, 678-686.

[25] Phillip Y, K. V., Schreiber G. (2012) Protein-binding dynamics imaged in a living cell, *Proc Natl Acad Sci U S A 109*, 1461-1466.

[26] Sukenik S, R. P., Gruebele M. (2017) Weak protein-protein interactions in live cells are quantified by cell-volume modulation, *Proc Natl Acad Sci U S A 114*, 6776-6781.

[27] Yang Y, C. S.-N., Yang F, Li X-Y, Feintuch A, Su X-C, Goldfarb D. (2020) In-cell destabilization of a homo-dimeric protein complex detected by DEER spectroscopy, *Proc Natl Acad Sci U S A 117*, 20566-20575.

[28] Kaur, A. D., S. . (2020) Recent applications of FRET-based multiplexed techniques, *Trends in Analytical Chemistry 123*, 115777.

[29] Kenworthy, A. K. (2001) Imaging protein-protein interactions using fluorescence resonance energy transfer microscopy, *Methods 24*, 289-296.

[30] Dye, B. T., Schell, K., Miller, D. J., & Ahlquist, P. . (2005) Detecting protein-protein interaction in live yeast by flow cytometry, *Cytometry Part A 63A*, 77-86.

[31] Oyama R, T. H., Yonezawa M, et al. . (2006) Protein-protein interaction analysis by C-terminally specific fluorescence labeling and fluorescence cross-correlation spectroscopy., *Nucleic Acids Res 34*.

[32] Sadaie, W. H., Y.; Matsuda, M.; Aoki, K. (2014) Quantitative In Vivo Fluorescence Cross-Correlation Analyses Highlight the Importance of Competitive Effects in the Regulation of Protein-Protein Interactions, *Molecular and Cellular Biology 34*, 3272-3290.

[33] Crowley, P. B., Kyne, C., and Monteith, W. B. (2012) Simple and inexpensive incorporation of ^{19}F-Tryptophan for protein NMR spectroscopy, *Chem. Commun. 48*, 10681-10683.

[34] Byeon, I.-J. L., Louis, J. M., and Gronenborn, A. M. (2003) A Protein Contortionist: Core Mutations of GB1 that Induce Dimerization and Domain Swapping, *Journal of Molecular Biology 333*, 141-152.

[35] Speer SL, G. A., Patteson JB, Ehrmann BM, Pielak GJ. (2019) Controlling and quantifying protein concentration in Escherichia coli *Protein Sci 28*, 1307-1311.

[36] Chu IT, S. S., Pielak GJ. (2020) Rheostatic control of protein expression using tuner cells, *Biochemistry 59*, 733-735.

[37] Davis, C. M., Gruebele, M., and Sukenik, S. (2018) How does solvation in the cell affect protein folding and binding?, *Curr. Opin. Struct. Biol. 48*, 23-29.

[38] Kyne, C., and Crowley, P. B. (2017) Short Arginine Motifs Drive Protein Stickiness in the Escherichia coli Cytoplasm, *Biochemistry 56*, 5026-5032.

[39] Powers, E., Powers, D., and Gierasch, L. (2012) FoldEco: A Model for Proteostasis in *E. coli*, *Cell Reports 1*, 265-276.

[40] Gronenborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P. T., and Clore, G. M. (1991) A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G., *Science 253*, 657-661.

[41] Byeon, I.-J. L., Louis, J. M., and Gronenborn, A. M. (2004) A Captured Folding Intermediate Involved in Dimerization and Domain-swapping of GB1, *J. Mol. Biol. 340*, 615-625.

[42] Nadeau, J. L. (2016) *Introduction to Experimental Biophysics: Biological Methods for Physical Scientists*, CRC Press.

[43] Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H. R., and Iwata, S. (2003) Structure and mechanism of the lactose permease of Escherichia coli, *Science 301*, 610-615.

[44] Taylor, P. J. (2005) Matrix effects: the Achilles heel of quantitative high-performance liquid chromatography-electrospray-tandem mass spectrometry, *Clin Biochem 38*, 328-334.

[45] Koch, A. R., B.; Button, D. . (1996) Deduction of the cell volume and mass from forward scatter intensity and of bacteria analyzed by flow cytometry, *J. Microbiol. Methods 27*, 49-61.

[46] Stock, J. B., Rauch, B., and Roseman, S. (1977) Periplasmic space in *Salmonella typhimurium* and *Escherichia coli*, *J. Biol. Chem. 252*, 7850-7861.

[47] Bouvier, T., Troussellier, M., Anzil, A., Courties, C., and Servais, P. (2001) Using light scatter signal to estimate bacterial biovolume by flow cytometry, *Cytometry 44*, 188-194.

[48] Bainer, R., Park, H., and Cluzel, P. (2003) A high-throughput capillary assay for bacterial chemotaxis, *J. Microbiol. Methods 55*, 315-319.

[49] Slade, K. M., Baker, R., Chua, M., Thompson, N. L., and Pielak, G. J. (2009) Effects of recombinant protein expression on green fluorescent protein diffusion in *Escherichia coli*, *Biochemistry 48*, 5083–5089.

[50] Tzur, A., Moore, J. K., Jorgensen, P., Shapiro, H. M., and Kirschner, M. W. (2011) Optimizing Optical Flow Cytometry for Cell Volume-Based Sorting and Analysis, *PLoS ONE 6*, e16053.

[51] Volkmer, B., and Heinemann, M. (2011) Condition-Dependent Cell Volume and Concentration of Escherichia coli to Facilitate Data Conversion for Systems Biology Modeling, *PLoS ONE 6*, e23126.

[52] Link, A. J., Robison, K., and Church, G. M. (1997) Comparing the predicted and observed properties of proteins encoded in the genome of *Escherichia coli* K-12, *Electrophoresis 18*, 1259-1313.

[53] Barnes, C. O., and Pielak, G. J. (2011) In-cell protein NMR and protein leakage, *Proteins 79*, 347-351.

[54] Burz, D. S., Dutta, K., Cowburn, D., and Shekhtman, A. (2006) Mapping structural interactions using in-cell NMR spectroscopy (STINT-NMR), *Nat. Methods 3*, 91-93.

[55] Burz, D. S., and Shekhtman, A. (2008) In-Cell Biochemistry Using NMR Spectroscopy, *PloS ONE 3*, e2571.

[56] Pielak, G. J., Li, C., Miklos, A. C., Schlesinger, A. P., Slade, K. M., Wang, G. F., and Zigoneanu, I. G. (2009) Protein nuclear magnetic resonance under physiological conditions, *Biochemistry 48*, 226-234.

[57] Majumder, S., DeMott, C. M., Burz, D. S., and Shekhtman, A. (2014) Using Singular Value Decomposition to Characterize Protein–Protein Interactions by In-cell NMR Spectroscopy, *ChemBioChem 15*, 929-933.

[58] Record, M. T., Jr., Courtenay, E. S., Cayley, D. S., and Guttman, H. J. (1998) Response of *E. coli* to osmotic stress: large changes in amounts of cytoplasmic solutes and water, *Trends Biochem. Sci. 23*, 143-148.

[59] Dedmon, M. M., Patel, C. N., Young, G. B., and Pielak, G. J. (2002) FlgM gains structure in living cells, *Proc. Natl. Acad. Sci. USA 99*, 12681-12684.

[60] Guseman, A. J., Perez Goncalves, G. M., Speer, S. L., Young, G. B., and Pielak, G. J. (2018) Protein shape modulates crowding effects, *Proc. Natl. Acad. Sci. USA 115*, 10965-10970.

[61] Jee, J., Byeon, I.-J. L., Louis, J. M., and Gronenborn, A. M. (2008) The point mutation A34F causes dimerization of GB1, *Proteins: Struct., Funct., Bioinf. 71*, 1420-1431.

[62] Barnes, C. O., and Pielak, G. J. (2011) In-cell protein NMR and protein leakage, *Proteins: Struct., Funct., Bioinf. 79*, 347-351.

[63] Speer, S. L., Guseman, A. J., Patteson, J. B., Ehrmann, B. M., and Pielak, G. J. (2019) Controlling and quantifying protein concentration in *Escherichia coli*, *Protein Sci 28*, 1307-1311.

[64] Sarkar, M., Li, C., and Pielak, G. J. (2013) Soft interactions and crowding, *Biophys. Rev. 5*, 187-194.

[65] Cong, Q., Anishchenko, I., Ovchinnikov, S., and Baker, D. (2019) Protein interaction networks revealed by proteome coevolution, *Science 365*, 185.

[66] van Holde, K. E. (1985) *Physical Biochemistry*, Prentice Hall.

[67] Guseman, A. J., and Pielak, G. J. (2020) Protein Stability and Weak Intracellular Interactions, In *In-cell NMR Spectroscopy: From Molecular Sciences to Cell Biology*, pp 188-206, Royal Society of Chemisty.

[68] Zimmerman, S. B., and Trach, S. O. (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli, *J Mol Biol 222*, 599-620.

[69] Theillet, F.-X., Binolfi, A., Frembgen-Kesner, T., Hingorani, K., Sarkar, M., Kyne, C., Li, C., Crowley, P., Gierasch, L., Pielak, G. J., Elcock, A., Gershenson, A., and Selenko, P. (2014) Physicochemical properties of cells and their effects on intrinsically disordered proteins (IDPs), *Chem. Rev. 114*, 6661-6714.

[70] Kozlowski, L. P. (2017) Proteome-pI: proteome isoelectric point database, *Nucleic Acids Res. 45*, 1112-1116.

[71] McConkey, E. H. (1982) Molecular evolution, intracellular organization, and the quinary structure of proteins, *Proc. Natl. Acad. Sci. U. S. A. 79*, 3236-3240.

[72] Pareek, V., Tian, H., Winograd, N., and Benkovic, S. J. (2020) Metabolomics and mass spectrometry imaging reveal channeled de novo purine synthesis in cells, *Science 368*, 283.

[73] Anfinsen, C. B. (1973) Principles that govern the folding of protein chains, *Science 181*, 223-230.

[74] Deeds, E. J., Ashenberg, O., Gerardin, J., and Shakhnovich, E. I. (2007) Robust protein protein interactions in crowded cellular environments, *Proc Natl Acad Sci U S A 104*, 14952-14957.

[75] You, X., Nguyen, A. W., Jabaiah, A., Sheff, M. A., Thorn, K. S., and Daugherty, P. S. (2006) Intracellular protein interaction mapping with FRET hybrids, *Proc. Natl. Acad. Sci. USA 103*, 18458-18463.

[76] Sudhaharan, T., Liu, P., Foo, Y. H., Bu, W., Lim, K. B., Wohland, T., and Ahmed, S. (2009) Determination of in vivo dissociation constant, KD, of Cdc42-effector complexes in live mammalian cells using single wavelength fluorescence cross-correlation spectroscopy, *J Biol Chem 284*, 13602-13609.

[77] Sukenik, S., Ren, P., and Gruebele, M. (2017) Weak protein–protein interactions in live cells are quantified by cell-volume modulation, *Proc. Natl. Acad. Sci. U. S. A. 114*, 6776-6781.

[78] Guin, D., and Gruebele, M. (2020) Chaperones Hsc70 and Hsp70 Bind to the Protein PGK Differently inside Living Cells, *J Phys Chem B 124*, 3629-3635.

[79] Yang, Y., Chen, S. N., Yang, F., Li, X. Y., Feintuch, A., Su, X. C., and Goldfarb, D. (2020) In-cell destabilization of a homodimeric protein complex detected by DEER spectroscopy, *Proc Natl Acad Sci U S A 117*, 20566-20575.

[80] Ye, Y., Liu, X., Zhang, Z., Wu, Q., Jiang, B., Jiang, L., Zhang, X., Liu, M., Pielak Gary, J., and Li, C. (2013) ^{19}F NMR Spectroscopy as a Probe of Cytoplasmic Viscosity and Weak Protein Interactions in Living Cells, *Chem. - Eur. J. 19*, 12705-12710.

[81] Aramini, J. M., Hamilton, K., Ma, L.-C., Swapna, G. V. T., Leonard, Paul G., Ladbury, John E., Krug, Robert M., and Montelione, Gaetano T. (2014) ^{19}F NMR Reveals Multiple Conformations at the Dimer Interface of the Nonstructural Protein 1 Effector Domain from influenza A Virus, *Structure 22*, 515-525.

[82] Guseman, A. J., and Pielak, G. J. (2017) Cosolute and Crowding Effects on a Side-By-Side Protein Dimer, *Biochemistry 56*, 971–976.

[83] Chu, I. T., Speer, S. L., and Pielak, G. J. (2020) Rheostatic Control of Protein Expression Using Tuner Cells, *Biochemistry 59*, 733-735.

[84] Speer, S. L., Guseman, A. J., Patteson, J. B., Ehrmann, B. M., and Pielak, G. J. (2019) Controlling and quantifying protein concentration in *Escherichia coli*, *Protein Science 28*, 1307-1311.

[85] Cicirelli, M. F., Robinson, K. R., and Smith, L. D. (1983) Internal pH of Xenopus oocytes: a study of the mechanism and role of pH changes during meiotic maturation, *Dev Biol 100*, 133-146.

[86] Ye, Y., Liu, X., Chen, Y., Xu, G., Wu, Q., Zhang, Z., Yao, C., Liu, M., and Li, C. (2015) Labeling strategy and signal broadening mechanism of Protein NMR spectroscopy in *Xenopus laevis* oocytes, *Chemistry 21*, 8686-8690.

[87] Parisis, N. (2012) *Xenopus laevis* as a Model-System, *Materials and Methods 2*.

[88] Lindman, S., Xue, W.-F., Szczepankiewicz, O., Bauer, M. C., Nilsson, H., and Linse, S. (2006) Salting the Charged Surface: pH and Salt Dependence of Protein G B1 Stability, *Biophysical Journal 90*, 2911-2921.

[89] Leeb, S., Sörensen, T., Yang, F., Mu, X., Oliveberg, M., and Danielsson, J. (2020) Diffusive protein interactions in human versus bacterial cells, *Current Research in Structural Biology 2*, 68-78.

[90] Mu, X., Choi, S., Lang, L., Mowray, D., Dokholyan, N. V., Danielsson, J., and Oliveberg, M. (2017) Physicochemical code for quinary protein interactions in Escherichia coli, *Proc Natl Acad Sci U S A 114*, E4556-E4563.

[91] Schavemaker, P. E., Śmigiel, W. M., and Poolman, B. (2017) Ribosome surface properties may impose limits on the nature of the cytoplasmic proteome, *eLife 6*, e30084.

[92] Ye, Y., Wu, Q., Zheng, W., Jiang, B., Pielak, G., Liu, M., and Li, C. (2019) Positively-Charged Tags Impede Protein Mobility in Cells as Quantified by ^{19}F NMR, *Journal of Physical Chemistry 123*, 4527-4533.

[93] Eaton, W. A. (2020) Hemoglobin S polymerization and sickle cell disease: A retrospective on the occasion of the 70th anniversary of Pauling's Science paper, *Am J Hematol 95*, 205-211.

[94] Mirceta, S., Signore, A. V., Burns, J. M., Cossins, A. R., Campbell, K. L., and Berenbrink, M. (2013) Evolution of Mammalian Diving Capacity Traced by Myoglobin Net Surface Charge, *Science 340*, 1234192.

[95] Cai, J., Townsend, J. P., Dodson, T. C., Heiney, P. A., and Sweeney, A. M. (2017) Eye patches: Protein assembly of index-gradient squid lenses, *Science 357*, 564.

[96] Smith, C. K., Withka, J. M., and Regan, L. (1994) A Thermodynamic Scale for the .beta.-Sheet Forming Tendencies of the Amino Acids, *Biochemistry 33*, 5510-5517.

[97] Gill, S. C., and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino-acid sequence data, *Anal. Biochem. 182*, 319-326.

[98] Li, C., Wang, G.-F., Wang, Y., Creager-Allen, R., Lutz, E. A., Scronce, H., Slade, K. M., Ruf, R. A., Mehl, R. A., and Pielak, G. J. (2010) Protein ^{19}F NMR in *Escherichia coli*, *J. Am. Chem. Soc. 132*, 321–327.

[99] Ye, Y., Liu, X., Xu, G., Liu, M., and Li, C. (2015) Direct Observation of Ca^{2+}-Induced Calmodulin Conformational Transitions in Intact *Xenopus laevis* Oocytes by ^{19}F NMR Spectroscopy, *Angew. Chem., Int. Ed. 54*, 5328-5330.